doispontos

Revista dos Departamentos de Filosofia
da Universidade Federal do Paraná e da Universidade Federal de São Carlos

Racionalidade e irracionalidade social

volume 22 número 3
dezembro de 2025

dois*pontos* é uma revista dos Departamentos de Filosofia da Universidade Federal do Paraná e da Universidade Federal de São Carlos vinculada aos Programas de Pós-graduação da Universidade Federal do Paraná e da Universidade Federal de São Carlos. Publica artigos de filosofia e de áreas afins com interesse filosófico e busca promover intercâmbio entre pesquisadores no Brasil e exterior.

ISSN: 2179–7412

endereço para correspondência||address for correspondence
Departamento de Filosofia da Universidade Federal do Paraná (UFPR)
R. Dr. Faivre, 405, 6° andar. CEP: 80060–140 – Curitiba, PR, Brasil.
(41) 3360–5098
Departamento de Filosofia da Universidade Federal de São Carlos (UFSCar) Rodovia Washington Luís, km 235. CEP: 13565–905 – São Carlos, SP, Brasil. (16) 3351–8366

endereços eletrônicos da *doispontos:*
revista2pontos@gmail.com
www.ser.ufpr.br/doispontos
www.filosofia.ufpr.br

doispontos:
Revista dos Departamentos de Filosofia da Universidade Federal do Paraná e da Universidade Federal de São Carlos

Índice

Editorial

Os ganhos advindos do processo de especialização na construção do conhecimento, bem como no desenvolvimento tecnológico e científico, podem ser significativamente dificultados, e por vezes até mesmo obstruídos, quando não são adequadamente absorvidos, compreendidos e reconhecidos pela sociedade. A produção científica não se esgota em seus próprios métodos e resultados; ela depende, em grande medida, de um ambiente social capaz de acolhê-la, interpretá-la e integrá-la às práticas coletivas. O processo pelo qual a ciência alcança relevância social é, portanto, permeado por atitudes tanto teóricas quanto práticas que não dizem respeito apenas aos cientistas, mas se estendem à sociedade como um todo.

Se é inegável que os conhecimentos científicos avançaram de modo extraordinário ao longo dos últimos dois séculos, também é verdade que, nas últimas décadas, e de maneira particularmente visível nos últimos vinte anos, testemunhamos o fortalecimento de formas variadas de anti-intelectualismo, bem como o surgimento de movimentos explicitamente hostis à autoridade epistêmica da ciência. A necessidade das Marchas pela Ciência, realizadas em diversas cidades do mundo a partir de 2017, constitui um sintoma eloquente desse cenário. Do mesmo modo, a pandemia de covid-19 evidenciou, em escala global, o quanto recomendações técnico-científicas puderam ser desconsideradas, relativizadas ou instrumentalizadas por interesses políticos e ideológicos. Esse quadro não se restringe a um evento específico: ele se prolonga nas resistências persistentes às evidências relativas ao aquecimento global e às mudanças climáticas, apesar do amplo consenso científico acerca de sua gravidade.

Diante desse contexto, torna-se urgente investigar as fontes, muitas vezes inerentes ao próprio intelecto humano, que tornam os indivíduos propensos a sucumbir a paixões, egoísmos, misticismos e fanatismos capazes de corromper e obscurecer o julgamento. O aspecto mais inquietante desse fenômeno talvez resida no fato de que tais tendências não se limitam àqueles desprovidos de instrução formal. Não raramente, observamos indivíduos com formação acadêmica abandonarem procedimentos críticos fundamentais e aderirem a posições marcadamente dogmáticas. Aqueles de quem se esperava uma postura mais alinhada aos valores da investigação científica por vezes contribuem, eles próprios, para a disseminação da confusão epistêmica.

Essa preocupação, evidentemente, não é nova na história da filosofia. Francis Bacon, um dos pais fundadores da ciência moderna, já alertava para os *idola*, os ídolos do intelecto, que inclinam o espírito humano ao erro. Ídolos da tribo, da caverna, do foro e do teatro designam, cada qual a seu modo, disposições estruturais que interferem na apreensão da verdade e exigem vigilância constante. A modernidade filosófica nasce, em parte, dessa consciência de que o progresso do conhecimento requer não apenas métodos adequados, mas também um trabalho crítico de depuração das ilusões que acompanham o exercício da razão.

Tampouco Kant deixou de reconhecer esse problema. Em diversos momentos de sua obra, e de maneira paradigmática na *Crítica da razão pura*, o filósofo tematizou as ilusões às quais a razão inevitavelmente se vê exposta. Longe de serem meros acidentes, tais ilusões emergem do próprio funcionamento da racionalidade quando esta ultrapassa os limites da experiência possível. A razão, por assim dizer, engendra suas próprias miragens. Reconhecer essa propensão não implica ceticismo, mas antes a necessidade de uma crítica permanente, capaz de delimitar o uso legítimo das nossas faculdades cognitivas.

Investigar as fontes da irracionalidade social constitui, assim, uma condição sine qua non para compreender por que muitos dos conhecimentos produzidos pela ciência, em seus mais diversos ramos, vêm sendo ideologicamente rejeitados, deformados ou seletivamente apropriados. Também ajuda a explicar por que profissionais com formação técnica altamente especializada (médicos, juristas, engenheiros, economistas, professores etc) por vezes abdicam do rigor que deu origem ao seu campo de conhecimento para aderir a narrativas simplificadoras ou identitárias, contribuindo para o adensamento de um ambiente público marcado pela desinformação e polarização ideológica.

Seria um equívoco, contudo, tratar esse processo como algo meramente artificial ou fabricado. Trata-se de um processo que combina disposições cognitivas e afetivas profundamente enraizadas junto com estruturas políticas, econômicas, sociais e midiáticas que atuam de forma relevante na amplificação da irracionalidade coletiva. Dessa forma, o ser humano parece oscilar entre duas tendências fundamentais: de um lado, o impulso e instituições que se direcionam a promoção do conhecimento, à explicação e à compreensão; de outro, forças e patologias sociais e institucionais que minam esse mesmo movimento, como o apego a crenças confortáveis, a busca por pertencimento, a resistência à revisão de convicções e a atração por narrativas que simplificam a complexidade do real.

Compreender essa tensão talvez seja um dos desafios centrais do nosso tempo. Pois se o desenvolvimento técnico-científico continuar a avançar em descompasso com a capacidade social de assimilá-lo criticamente, corremos o risco de assistir a uma situação paradoxal: jamais produzimos tanto conhecimento, e ainda assim ele pode se tornar socialmente inócuo. Combater esse cenário exige mais do que a simples reafirmação da autoridade da ciência. Exige compreender as raízes das tendências irracionais que atravessam tanto a vida individual quanto a coletiva. Somente a partir de um diagnóstico mais preciso será possível discutir estratégias de enfrentamento ou, ao menos, desenvolver formas de precaução intelectual e institucional. Trata-se não apenas de defender conteúdos científicos específicos, mas de fortalecer as condições culturais que tornam possível o exercício público da razão.

Foi precisamente o reconhecimento da complexidade desse problema que orientou a concepção deste volume da *Revista Dois Pontos*. Ao reunir contribuições provenientes de diferentes tradições filosóficas, abordagens teóricas e perspectivas metodológicas, buscamos lançar luz sobre as múltiplas dimensões da irracionalidade social. Mais do que oferecer respostas definitivas, os textos aqui reunidos pretendem mapear o terreno, explicitar tensões e abrir caminhos para investigações futuras.

Reconhecer e analisar a existência dessa tensão entre racionalidade e irracionalidade não é apenas um exercício teórico; trata-se de uma tarefa com implicações profundamente práticas. Em um momento histórico no qual as fronteiras entre conhecimento e opinião parecem cada vez mais difusas, reafirmar o valor da crítica, da argumentação e do exame rigoroso torna-se um imperativo filosófico e civilizacional. Este volume convida, portanto, o leitor a refletir sobre um problema que, longe de ser periférico, toca o próprio destino das sociedades contemporâneas. Pois, se é verdade que o ser humano possui uma inclinação natural para conhecer, não é menos verdadeiro que carrega consigo tendências que ameaçam continuamente esse projeto. Tornar visível essa ambivalência talvez seja o primeiro passo para que possamos enfrentá-la com a lucidez que ela exige.

Prof. Dr. Joel Thiago Klein (UFPR)

Kant's wager?: Betting as a touchstone of subjective conviction

Uma aposta de Kant?: A aposta como pedra de toque da convicção subjetiva

Jean-Christophe Merle
Universität Vechta
jean-christophe.merle@uni-vechta.de

Abstract: In the *Third Section. On having an opinion, knowing, and believing* of "The Canon of Pure Reason" of the *Critique of Pure Reason,* Kant writes: "Taking something to be true, or the subjective validity of judgment, has the following three stages in relation to conviction (which at the same time is valid objectively): having an opinion, believing and knowing [*Meinen, Glauben, Wissen*] (A822/B850). This article attempts to show that P. Guyer and A.W. Wood's translation of Glaube is inadequate. The translation rests on false premises and leads to a momentous misunderstanding of what is at stake in "The Canon of Pure Reason" as well as its entire thesis and argument. This article offers a new interpretation of the Third Section, in which rational faith is put at the same level of certainty as knowledge.

Keywords: belief; canon; faith; Kant; knowledge; opinion.

Resumo: Na *Terceira Seção: Da opinião, da ciência e da fé* de "O cânone da razão pura" da *Crítica da razão pura*, Kant escreve: "A crença ou a validade subjetiva do juízo, relativamente à convicção (que tem ao mesmo tempo uma validade objetiva), apresenta os três graus seguintes: opinião, fé e ciência [*Meinen, Glauben, Wissen]*" (A822/B850). O presente artigo tenta demonstrar que a tradução de P. Guyer e A.W. Wood para Glaube é inadequada. A tradução repousa sobre falsas premissas e nos conduz a um grave equívoco na compreensão daquilo que está em jogo em "O cânone da razão pura", assim como de sua tese e argumento como um todo. Este artigo apresenta uma nova interpretação da Terceira Seção, na qual a fé racional é posta no mesmo nível da certeza como conhecimento.

Palavras-chave: crença; cânone; fé; Kant; conhecimento; opinião.

Recebido em 10 de setembro de 2025. Aceito em 11 de dezembro de 2025.

1) Four problematic assumptions about Glauben and Wissen in Kant

In "The Canon of Pure Reason" of the Critique of Pure Reason, Kant famously asserts:

> Taking something to be true [*das Fürwahrhalten*], or the subjective validity of judgment, has the following three stages [*Stufen*] in relation to conviction (which at the same time is valid objectively): having an opinion, believing and knowing [*Meinen, Glauben, Wissen*] (A822/B850).

Since, depending on the context, the German word Glaube may be translated either as "belief" or as "faith", I provide in this quotation of the Cambridge edition by Paul Guyer and Allen W. Wood the German original in square brackets. What is at stake with the adequacy of the translation of this word in the "Third Section. On having an opinion, knowing, and believing" is nothing less than a correct understanding of that whole section, and of the "The Canon of Pure Reason" itself.

Commentators usually make the following four assumptions about this passage.

First, they assume that Glauben refers both to "doctrinal beliefs" (*den doktrinalen Glauben,* is plural in the original too) (A825/B853) and to "moral belief" (*dem moralischen Glauben*, is singular in the original too) (A828/B856) in the following passages:

> [...] there is in merely theoretical judgments an analogue of practical judgments, where taking them to be true is aptly described by the word belief [*Glauben*], and which we can call doctrinal beliefs [*Glauben*] (KrV, A 825/B 853).
> But there is something unstable about merely doctrinal belief [*doktrinale Glaube*]; one is often put off from its difficulties that come up in speculation, although, to be sure, one inexorably returns to it again. It is entirely otherwise in the case of moral belief. For it is absolutely necessary that something must happen, namely, that I fulfill the moral law in all points (KrV A 828/B 856).

Second, they assume that the meaning of Glauben is "belief".

Third, they assume that knowledge [*Wissen*] has nothing in common with *Glauben*. The evidence usually provided for this third assumption is found in the following passage:

> Having an opinion is taking something to be true with the consciousness that it is subjectively as well as objectively insufficient. If taking something to be true is only subjectively sufficient and is at the same time held to be objectively insufficient, then it is called believing [*Glauben*]. Finally, when taking something to be true is both subjectively and objectively sufficient it is called knowing [*Glauben*]. Subjective sufficiency is called conviction [*Überzeugung*] (for myself), objective sufficiency, certainty [*Gewißheit*] (for everyone) (A822/B850).

Fourth, they assume that both "doctrinal beliefs" and "moral belief" are objectively insufficient, and thus are not knowledge [*Wissen*].

These four assumptions are highly problematic on at least three points.

First, they do not provide any clear information about what is exactly the object of "Wissen" in "The Canon of Pure Reason", and more precisely in its "Third section. On having opinions, knowing [*Wissen*], and believing [*Glauben*]".

Second, they do not give any clear information about the way in which the three "stages" [*Stufen*] – or, more literally, "levels" – are really ordered. On this point, the "Third section" seems to provide pieces of information that are hardly compatible with one another, if one makes the assumptions mentioned above. Indeed, on the one hand, Kant mentions: "three stages [*Stufen*] [...]: having an opinion, believing and knowing [*Meinen, Glauben, Wissen*]" (A 822/B 850), while, on the other hand, the title of the whole section is "On having opinions, knowing [Wissen], and believing [*Glauben*]" (A 820/B 848). Now, whereas in both cases opinion [*Meinen*] continues to be the first "stage", Wissen and Glauben are reversed in order. Thus, if one assumes consistency,

one should either read in the order of the former quotation not three "stages" or Glauben in the latter quotation does not mean the same as Glauben in the former quotation. I will argue for the latter of the two options.

In fact, this problem of consistency brings the focus back to the issue of the adequate translation of *Glaube*. In this regard, there are two equivocations: (i) *Glaube* may mean either "belief" or "faith"; (ii) *Glaube* may refer either to doctrinal beliefs [*doktrinale Glauben*] or to moral faith [*moralischer Glaube*].

2) Offering a consistent and adequate translation and interpretation

Concluding from the problems of consistency mentioned above that there is a lack of consistency in Kant's famous passage on *Glaube* and *Wissen* should be our last resort solution. Therefore, I offer the following consistent interpretation, which rests on a suggestion for adequately translating *Glaube*.

Let us begin with the second "stage" mentioned above. Since this "second" stage, i.e. moralischer Glaube, is not related to any existing state of the world, unlike opinion [*Meinen*] and knowledge [*Wissen*], but, instead, to hope, I suggest distinguishing "doctrinal beliefs" from "moral faith". Thus, in this quotation, one ought to translate Glauben as belief, and to understand it as doctrinal beliefs. Since "doctrinal beliefs" are dismissed by Kant as speculative and "unstable", they cannot belong to the last stage or level. Thus, I suggest modifying the translation of one of the two quotations or making it more precise: "three stages" [*Stufen*] [...]: having an opinion, [*doctrinal*] beliefs and knowing [*Meinen, Glauben, Wissen*]." Thus, I will refer the other term of the equivocation mentioned above, i.e. "moral faith", to the other quotation: "On having opinions, knowing, and [*moral*] faith". Here, one may put knowing and moral faith on the same level, not only because there is no explicit conceptual order in the title of the section, rather, only a stylistic order between both, but also because of the following: knowing [*Wissen*] is both subjectively and objectively sufficient, and thus knowledge is considered by Kant as certain. Yet, in Kant's view, moral faith is certain too, so that knowing and faith have something in common that puts them at the same level, namely, certainty. This is shown clearly in the following passage:

> Of course, no one will be able to boast that he knows that there is a God and a future life [...]. All knowing (if it concerns an object of reason alone) can be communicated [...]. No, the conviction is not logical, but moral certainty, and, since it depends on subjective grounds (of moral disposition) I must not even say 'It is morally certain that there is a God,', etc., but rather 'I am morally certain' etc (KrV, A 829/B 857).

Admittedly, Kant also asserts: "Subjective sufficiency is called conviction [*Überzeugung*] (for myself), objective sufficiency, certainty [*Gewißheit*] (for everyone)." (KrV, A 823/B 850) Yet, there is no impossibility for moral certainty to be an objective certainty too. Indeed, one should conceptually distinguish between the communicability of the common judgment ("[...] the judgement of every understanding must agree", KrV, A 820/B 848) and the "common ground" (A 820/B 848) of this universal agreement. The common ground is the basis for communicability. Therefore, there is no incompatibility for conviction for myself and certainty for everyone to share the same common ground.

The reason why this point is frequently overlooked may lie in an assumption made by most of the commentators, and explicitly formulated by Thomas Höwing. He considers the order of the three Stufen as being guided by an "epistemic ideal" (HÖWING, 2018, p. 1249). Yet, this assumption is

incompatible with the object of "The Canon of Pure Reason", since its object is beyond the reach of cognition, and, hence, beyond the reach of epistemology too. Kant emphasizes:

> […] in a canon of pure reason we are concerned with only two questions that pertain to the practical interest of pure reason, and with regard to which a canon of its use must be possible, namely: Is there a God? Is there a future life? (KrV, A 803/B 831)

Instead of being related to an epistemic ideal, the "three stages [*Stufen*] in relation to conviction" (KrV, A 822/B 850) refer to a conviction concerning the existence of God and of a future life. Now, this conviction concerning the existence of God and of a future life is considered by Kant as being "valid objectively" (KrV, A 822/B 850), i.e., having "moral certainty" (A 829/B 857).

Now, Kant seems to mention two touchstones of this conviction. According to Joseph Trullinger, there are "two touchstones: communicability and betting" (TRULLINGER, 2013, p. 382). Yet, in my view, the designation of "communicability" for one of the touchstones is not appropriate. Indeed, mere communicability, on the one hand, and common judgment or common ground, on the other hand, are not exactly the same, as Habermas rightly observes:

> Already in the Critique of Pure Reason, Kant had ascribed the function of a pragmatic test of truth to the public consensus arrived at by those engaged in rational critical debate with one another: The touchstone whereby we decide whether our holding a thing to be true is conviction or mere persuasion is therefore external, namely the possibility of communicating it and of finding it to be valid for all human reason.' This agreement of all empirical consciousnesses, brought about in the public sphere, corresponds to the intelligible unity of transcendental consciousness (HABERMAS, 1989, 107 f.).

Accordingly, the two touchstones are accepting to or abstaining from betting and, more exactly, common judgment, which is at the basis of both Trullinger's alleged touchstone of communicability and agreement (consensual communication). Now, Trullinger affirms the existence of a "tension between the two touchstones" (TRULLINGER, 2013, p. 394). Yet, although there are two touchstones, there is, in my view, no evidence of any tension. On the contrary, as I will now try to show, there is a commonality of these two touchstones, or, to say it more precisely, they have a common ground.

In the "Third section" of "The Canon", there are two explicit passages about betting: one related to an unspecified issue and the second one, related to an issue of doctrinal belief. Both of them deal with betting as a criterion for distinguishing opinion and persuasion, on the one hand, from belief (second level) and conviction, on the other hand. Yet, in my view, there are two further passages on bets, although they are merely implicit ones. Although neither of them concern a positive bet, both of them concern the abstention from betting. Both passages pertain to faith, located at the same third "stage" as knowledge [*Wissen*].

Let us have a look at the first explicit passage about betting:

> The usual touchstone of whether what someone asserts is mere persuasion or at least subjective conviction, i.e. firm belief [festes Glauben], is betting [Wetten]. Often someone pronounces his propositions with such confident and inflexible defiance [Trotz] that he seems to have entirely laid aside all concern for error. A bet disconcerts him [macht ihn stutzig]. Sometimes he reveals that he is persuaded enough for one ducat but not for ten. For he would happily bet one, but at ten he suddenly becomes aware of what he had not previously noticed, namely that it is quite possible that he has erred. If we entertain the thought that we would wager the happiness of our whole life on something, our triumphant judgement would quickly disappear, we would become timid [schüchtern] and we would suddenly discover that our belief does not extend so far. Thus pragmatic belief has only a degree, which can be large or small according to the difference of the interest that is at stake (KrV, A 825f./B 853f.).

This passage addresses what current research in epistemology calls respectively cognitive virtues and cognitive vices, or "vices of the mind" (see QUASSAM, 2019). In this passage, cognitive vices are mentioned in the expressions "confident and inflexible defiance" and "triumphant", and cognitive virtues in the consideration "that it is quite possible that he has erred". Betting is clearly presented by Kant as a remedy against this vice, i.e., as a way to awaken doubts, discarding ("disappear") a passion in favor of a more important interest, and, therefore, paying closer attention to the issue at stake. Indeed, according to Kant, betting manifests two major cognitive virtues, namely modesty and firmness: "The expression of belief *[Glaubens] i*s in such cases an expression of modesty from an objective point of view, but at the same time of the firmness of confidence in a subjective one (KrV, A 827/B 855).

In this passage, Glaube is an expression of modesty because, from a theoretical point of view, it does not even claim to be a hypothesis, as Kant makes clear in the remaining part of the same paragraph. Indeed, it raises no cognitive claim at all. In order not to misinterpret this passage, one should not confuse modesty with timidity by which Kant understands the following character trait: "[…] timidity [*Blödigkeit*] [is] a kind of concern [*Besorgniß]* and shyness [*Schüchternheit*] not to appear favorably in the eyes of others." (Anth, AA 7: 257f.) In fact, passing the test of betting does not reveal any weakness of character such as timidity, but rather demonstrates firmness, or at least more firmness. Now, what is a touchstone, and what does firmness in the passage consist in?

Suppose that proposition p is "x firmly holds proposition r for true", proposition q "x bets on the truth of proposition r", and let us add the assumption that one can empirically check whether a is true. In this passage, Kant argues in the following way:

p -> q

¬q -> ¬p (contraposition)

Thus if ¬q (i.e., in the case of x's refusal to bet on r), then ¬p means that there is no subjective sufficiency, and that x holding r for true, that is p, was x's mere opinion. Before the test, one cannot know whether x has a conviction or a mere opinion.

At first sight, firmness consists in the absence of doubt, even considering the damage occurring in the case in which one's assertion would appear to be wrong. Now, in this passage, firmness is merely relative, as Kant observes: "Thus pragmatic belief has only a degree, which can be large or small according to the difference of the interest that is at stake." (KrV, A 825f./B 853f.) What is at stake? Respectively one ducat (i.e., 3,5 grams of gold), ten ducats (35 grams of gold), and the happiness of our whole life. The latter of these stakes is the supreme test or degree of firmness.

About this test, let us make the following observation:

[(p ->q) ^q] -/> p.

In other words, this test, if x passes it, does not positively prove that x has the belief that r is true. The test can only negatively prove, if x fails the test, that x does not have the belief that r is true. The second explicit passage on betting shows this point clearly.

In the second explicit passage about betting, Kant explains:

> Since, however, even though we might not be able to undertake anything in relation to an object, and taking something to be true is therefore merely theoretical, in many cases we can still conceive and imagine an undertaking for which we would suppose ourselves to have sufficient grounds if there were a means for arriving at certainty [*Gewißheit*] about the matter, thus there is in merely theoretical judgments an analogue of practical judgments, where taking them to be true is aptly described by the word belief, and which we can call doctrinal beliefs. If it were possible to settle by any sort of experience whether there are inhabitants of at least some [wenigstens in irgend einem] of the planets that we see, I might well bet everything that I have on it. Hence I say that it is not merely an opinion but a strong belief (on the correctness of which I would wager many advantages in life) that there are also inhabitants in other worlds" (KrV, A825/B853).

From all we know about Kant's moral theory, the existence of "inhabitants in other worlds" – that is, of rational beings in other worlds – is likely to be Kant's own wish. Support for this is found in his Doctrine of Right, where he emphasizes:

> It is possible for me [i.e. for a rational being] to have any external object of my choice as mine, that is, a maxim by which, if it were to become a law, an object of choice would in itself (objectively) have to belong to no one (res nullius) is contrary to right (RL, AA 6: 250).

Indeed, if other worlds exist, which we neither know of nor can have access to, the existence of rational beings in those worlds that possess objects of those worlds as objects of their choice adds objects of reasons to those objects of reason already existing in our world. Now, since Kant cannot experience those worlds, him putting money on their existence would not be a bet, because in a bet one has partial cognition, but, instead, a gamble, because in a gamble there is complete ignorance. Thus,

> [...] there is something unstable [*etwas Wankendes*] about merely doctrinal belief [*doktrinale Glaube*]; one is often put off from it by difficulties that come up in speculation, although, to be sure, one inexorably returns to it again (KrV, A827/B855).

Let us now examine the two passages that refer merely implicitly to betting.

The first implicit passage about betting exposes a case in which changing one's holding-for-true is not possible:

> But there is something unstable about merely doctrinal belief [*doktrinale Glaube*] [...]. It is entirely otherwise in the case of moral belief [*moralischen Glauben*]. For there it is absolutely necessary that something must happen, namely, that I fulfill the moral law in all points. The end here is inescapably fixed, and according to all my insight there is possible only a single condition under which this end is consistent with all ends together and thereby has practical validity, namely, that there be a God and a future world; I also know with complete certainty that no one else knows of any other conditions that leads to this same unity of ends under the moral law. But since the moral precept is thus at the same time my maxim (as reason commands that it ought to be), I will inexorably believe in the existence of God and a future life, and I am sure that nothing can make these beliefs unstable, since my moral principles themselves, which I cannot renounce without becoming contemptible [*verabscheuungswürdig*] in my own eyes, would thereby be subverted (KrV, A 828/B 856).

Notice that by holding the existence of God and of a future world for true, I do not abandon modesty, since I still do not claim to have cognition. Thus, this passage is about Glaube as faith, not as belief, unlike in Paul Guyer and Allen W. Wood's translation. The following will explain how abstention from this implicit bet is irrational.

In order to present the reason for its irrationality, let us logically analyze this passage above.

Suppose that the proposition p is "I fulfill the moral law in all points", the proposition q "I bet (i.e. I have faith) in the existence of God and of a future world" – a faith which is the postulates of practical reason. Notice that one cannot empirically check whether p or q is true. Now, the argument is:

p

p ->q

¬q -> ¬p

¬q

¬p

In other words, this bet is self-contradicting. It contradicts the assertion of me being a rational being, which is a presupposition of my fulfillment of the moral law in all points. Indeed, ¬p means that it appears to me – one who thought that "I fulfill the moral law in all points", which I consider to be the inescapably right end – that I don't do that. I am revealed to myself as being verabscheuungswürdig, i.e., completely repulsive to myself (*Abscheu* is repulsive in its very definition), and this repulsion is not momentary, but constitutive of me as a rational being. Thus, betting in this case demonstrates stable firmness, that is, certainty.

Now, what is at stake in this implicit bet is me being a rational being, i.e., what is at stake is my own self. If I don't bet, I lose nothing less than my own self.

Finally, let us examine the second implicit passage about betting:

> If we […] assume someone who would be entirely indifferent in regard to moral questions, then the question that is propounded by reason becomes merely a problem for speculation, and in that case it can be supported with strong grounds from analogy but not with grounds to which even the most obstinate skepticism must yield. But no human being is free of all interest in these questions. For although he might be separated from the moral interest by the absence of all good dispositions, yet even in this case there is enough left to make him fear a divine existence and a future [life: *Leben*; a word missing in the Cambridge edition]. For to this end, nothing more is required than that he at least cannot pretend to any certainty [*Gewißheit*] there is no such being and no future life, which would have to be proved through reason alone and thus apodictically since he would have to establish them to be impossible, which certainly no rational human can undertake to do. That would be a negative belief [*ein negativer Glaube*], which, to be sure, would not produce morality and good dispositions, but would still produce the analogue of them, namely it could powerfully restrain the outbreak of evil dispositions (A 829f./B 857f.).

Let us analyze the part of the accessory sentence "[…] that he at least cannot pretend […] no rational human can undertake to do." from the point of view of modal logic, supposing that r is the proposition "x fears to be punished by God in a future life for his or her immoral deeds":

[](◊(r))

Let us now suppose that p is the proposition "x is amoral", that q is the proposition "x is confident in the absence of any God and of any future life". Let us notice that q is an assertion that cannot be empirically checked, so that it is a bet.

What does the irrationality of abstaining from betting in this case consist in? It is self-contradiction, which can be demonstrated as follows:

p

p ->q

¬q-> ¬p

r -> ¬q

r

¬q

¬p

We should understand the affirmation that "there is enough left to make him fear a divine existence and a future life" having in mind the following characterization of fear: "Anxiety, anguish, horror, and terror are degrees of fear, that is, degrees of aversion to danger." (Anth, AA 7: 256) Thus, in this case too, the abstention from betting generates repulsion in the rational being.

There are at least two commonalities between the respective abstentions of the two cases of implicit bet. First, there is a strong repulsion, which is displayed by the expressions "contemptible [*verabscheuungswürdig*]" and "there is enough left to make him fear a divine existence and a future life". Second, both cases are about an infinite loss, which makes them radically different from the famous wager presented in Blaise Pascal's Pensées, which is about an infinite gain. Indeed, in Pascal's wager, if God exists, he will infinitely reward those (and only those) who believe in him.

3) A new interpretation of the "stages" and its consequences

In all four passages about betting, the certainty of holding-to-be-true is based on the correspondence between the bet and the truth. In fact, Kant affirms:

> Truth, however, rests upon agreement with the object, with regard to which, consequently, the judgments of every understanding must [*müssen*] agree (consentientia uni tertio, consentiunt inter se). The touchstone [*Probierstein*] of whether taking something to be true is conviction [*Überzeugung*] or mere persuasion [*Überredung*] is, therefore, externally, the possibility of communicating it and finding it to be valid for the reason of every human being to take it to be true; for in that case there is at least a presumption [*Vermutung*] that the ground of the agreement of all judgments, regardless of the difference among the subjects, rests of the common ground, namely the object, with which they therefore all agree and through which the truth of the judgement is proved" (KrV, A 820/B 848).

In empirical issues, Kant's correspondence theory of truth refers to an object of experience. In moral issues, the correspondence is to what is constitutive of any rational being (which, in this case, is the "object" to which faith refers), so that it can be found "to be valid for the reason of every human being to take it to be true". Admittedly, one should point out that even if x necessarily has faith that p, it does not prove that the proposition p is true. Nevertheless, it proves that the proposition p is not less certain than in the case in which x had knowledge that p [*Wissen*]. Providing certainty, faith is a stable conviction, which is what Kant taught to his students:

> Belief [*Glaube*, rather: faith] is firm, then, when it leads a rational man to neglect the advantages of his life for his belief [*Glauben*, rather: faith]. He who is moved by duty and hope [*Hoffnung*] combined to renounce all these advantages believes [*glaubt*, rather: has faith] and is convinced [überzeugt]. In regard to its effect on the subject, this holding-to-be-true [*Fürwahrhalten*] will not yield to the highest certainty, and practical conviction is the strongest possible [*höchst mögliche*]. This practical conviction [*praktische Überzeugung*] can fall on certain propositions, and these are then morally certain propositions [*moralisch gewiße Sätze*]. These are the ground of all morality, and they agree with our greatest conscientiousness [*Gewissenhaftigkeit*], if we live according to them and thus coordinate our actions to them (V-Lo/Wiener, AA 24: 855).

From this arises a new interpretation of the three "stages" mentioned in the passages that I quoted at the beginning of this essay:

> Taking something to be true [*das Fürwahrhalten*], or the subjective validity of judgment, has the following three stages [*Stufen*] in relation to conviction (which at the same time is valid objectively): having an opinion, believing and knowing [*Meinen, Glauben, Wissen*] (A 822/B 850).

The three Stufen, for which a better translation would be "three levels", on a graduated scale of increasing certainty are:

(1) having an opinion;

(2) believing [*Glaube*], in the sense of doctrinal beliefs [*Doktrinale Glauben*];

(3) Knowledge [*Wissen*] and faith [*Glaube*], because both have certainty, either theoretical (in the case of *Wissen*) or practical (in the case of *moralischer Glaube*)

On this basis, one can explain two passages of Kant's writings that are wrongly translated, and that are as famous as they are frequently misunderstood. The first one belongs to the Critique of Pure Reason:

> [...] I cannot even assume [*annehmen*] God, freedom and immortality for the sake of the necessary practical use of my reason unless I simultaneously deprive speculative reason of its pretension to extravagant insights; because in order to attain to such insights, speculative reason would have to help itself to principles that in fact reach only to objects of possible experience, and which, if they were to be applied to what cannot be an object of experience, then they would always actually transform it into an appearance, and thus declare all practical extension of pure reason to be impossible. Thus, I had to deny [aufheben] knowledge [*Wissen*] in order to make room for faith [*Glauben*]; and the dogmatism of metaphysics [...] is the true source of all unbelief [*Unglaubens*, rather unfaith] conflicting with morality, which unbelief is always very dogmatic (KrV, B XXX).

The verb "aufheben" should be understood as also meaning, at the same time, "aufbewahren", i.e., "sorgsam hüten", keeping, retaining, preserving. In this case, "aufheben" means that one has to deny the "extravagant" claims of speculative reason (in classical metaphysics) about God, freedom and immortality, and to preserve (theoretical) certainty in empirical matters (i.e. one has to limit theoretical certainty to empirical matters), in order to remove the obstacle to (moral) faith, safeguarding moral faith from damages (indirectly) inflicted by the "extravagant" claims of speculative reason.

The second passage of Kant's writings that should be reinterpreted in the light of this new interpretation is a famous passage of the Critique of the Power of Judgment:

> We can thus assume a righteous man (like Spinoza) who takes himself to be firmly convinced [überredet, rather: persuaded] that there is no God and (since with regard to the object of morality it has a similar consequence) there is also no future life: how would he judge his own inner purposive determination by the moral law, which he actively honors? He does not demand any advantage for himself from his conformity to this law, whether in this or in another world [...]. But his effort is limited; and from nature he can, to be sure, expect some contingent assistance here and there, but never a lawlike [gesetzmäßige] agreement in accordance with constant rules (like his internal maxims are and must be) with the ends to act in behalf of which he still feels himself bound and impelled [*zu dem Zwecke erwarten, welchen zu bewirken er sich doch verbunden und angetrieben fühlt*]. Deceit, violence, and envy will always surround him, even though he is himself honest, peaceable, and benevolent; and the righteous ones besides himself that he will still encounter will, in spite of all their worthiness to be happy, nevertheless be subject by nature, which pays no attention to that, to all the evils of poverty, illness, and untimely death, just like all the other animals on earth, and will always remain thus [this: *es*] until one wide grave engulfs them all together (whether honest or dishonest, it makes no difference here and flings them, who where capable of having believed themselves to be the final end of creation, back into the abyss of purposeless chaos of matter from which they were drawn. – The end, therefore, which this well-intentioned person had and should have had before his eyes in his conformity to the moral law,

> he would certainly have to give up as impossible; or, if he would remain attached to the appeal of his moral inner vocation, [...] then he must assume the existence of a moral author of the world, i.e. of God, from a practical point of view, i.e., in order to form a concept of at least the possibility of the final end that is described to him by morality – which he very well can do, since it is at least not self-contradictory" (KU, AA 5: 452).

In Spinoza's case, the touchstone of accepting to bet or abstaining from betting would demonstrate either (i) that Spinoza really is a righteous man, in which case it would be revealed to him that he necessarily has to renounce his atheism, or (ii) that he is amoral. (i) and (ii) are incompatible with one another. Compatibility with Kant's writings and consistency are two criteria which I hope to have met in the present interpretation of "Meinen, Glauben, Wissen".

References

HABERMAS, J. 1989. *The structural transformation of the public sphere (1962)*. Trans. Thomas Burger & Frederick Lawrence. London: Polity.

HÖWING, T. 2018. Zur Vollständigkeit von Kants Unterscheidung zwischen Meinen, Glauben und Wissen. In: WAIBEL, V. L.; RUFFING, M.; WAGNER, D. (Eds.). *Natur und Freiheit*: Akten des XII. Internationalen Kant-Kongresses. Berlin: De Gruyter.

KANT, I. 1900-. *Gesammelte Schriften*. Königlich Preussische Akademie der Wissenschaften (and successors) (Eds.). Berlin: de Gruyter (and predecessors).

QUASSAM, Q. 2019. *Vices of the mind*: from the intellectual to the political. Oxford: Oxford University Press.

TRULLINGER, J. 2013. *Kant's two touchstones for conviction*: the incommunicable dimension of moral faith. The Review of Metaphysics, Washington D.C., v. 67, n. 2, p. 369-403, dec 2013.

Between control and trust: paradigms of early modern and Enlightenment optimism in the thought of Kant and Spener and their continuity in contemporary thought

Entre controle e confiança: paradigmas do otimismo no início da modernidade e no Iluminismo no pensamento de Kant e Spener e sua continuidade no pensamento contemporâneo

Anna Szyrwinska-Hörig
Universität Vechta
anna.szyrwinska@uni-vechta.de

Abstract: The article addresses the question of the origins of optimism in the perception of the future and the role that the sense of being in control of reality plays in its formation. This article analyses two models of optimism present in modern and Enlightenment thought: those of Ph.J. Spener and I. Kant. It considers the factors that contribute to optimism, including a sense of control over one's fate as a result of independence from sub-natural factors (such as God) in shaping the future and a sense of belief that reality is developing in an optimal way.

Keywords: control; Enlightenment; fate; optimism; pessimism; trust.

Resumo: O presente artigo aborda a questão das origens do otimismo na percepção de futuro e o papel que a ideia de controle da realidade desempenha em sua formação. O artigo analisa dois modelos de otimismo presentes no pensamento moderno e iluminista: os de Ph.J. Spener e I. Kant. O texto examina os fatores que contribuem para o otimismo, incluindo uma ideia de controle sobre o próprio destino como resultado da independência de fatores subnaturais (como Deus) na conformação do futuro e a crença de que a realidade está se desenvolvendo da melhor forma possível.

Palavras-chave: controle; Iluminismo; destino; otimismo; pessimismo; confiança.

Recebido em 10 de setembro de 2025. Aceito em 11 de dezembro de 2025.

The future can be approached in at least the two following and influential ways: optimistically or pessimistically. This raises the question of the source of optimism and pessimism. This paper will focus in particular on the question of optimism: What factors contribute to the development of a positive attitude in the individual and such expectations in society as a whole? In trying to answer these questions, let us look at the way of thinking that characterized the Enlightenment, an era in which an optimistic outlook on the future of humanity was one of its defining characteristics. Other typical features of the Enlightenment, including the conviction about the power of rationality, the belief that reasonableness is a universal quality and the demand for the construction of a universal and communicable system of knowledge, were shaped by the assumption that the future of humanity could potentially be improved. Optimism served both as the foundation for the development of other Enlightenment beliefs and as a driving force in the pursuit of Enlightenment objectives, self-reinforcing as a core tenet of the Enlightenment ideology. An intriguing question definitely worth looking into is the question of the ultimate source of the Enlightenment optimism. From what source did this optimistic belief emerge at that time? It is particularly captivating given that, at first glance, it appears to have no empirical justification. The evidence of human progress, for instance in the field of science, may suggest that further discoveries will be made but it does not provide sufficient grounds for an unquestioning belief in a better future.

The following article aims to reflect on the emergence of optimism as a worldview and examine the underlying rationality for this shift. This paper will focus on two kinds of optimism that may be observed in early modern theology and Enlightenment philosophy. It will compare notions concerning the predictions of the future developed by two classical thinkers, Philipp Jacob Spener (1635-1705) and Immanuel Kant (1724-1804). This choice is not accidental. Spener, who was not a representative of the Enlightenment but a forerunner of several Enlightenment ideas, is one of the very few theologians whose conception attracted Kant's attention. It would be inaccurate to claim that Spener had a significant influence on Kant, but there are sufficient parallels between their systems suggesting that they belonged to the same anthropological paradigm. In the context of the investigated topic, there is one striking similarity between Kant and Spener that is worthy of particular mention. Both Spener and Kant were strongly interested in the idea of moral progress and believed that such progress was possible. Their conviction about the real possibility of enhancement of human nature determined their belief that due to moral progress, a better future for the whole of humanity may be achieved. However, there is a crucial difference between Spener's and Kant's optimism concerning their visions of the future. Namely, Spener was convinced that the ultimate shape of the future depends on God's support and that no moral progress is achievable without God's grace. Kant, however, intended to develop a philosophical system in which the position of God in the traditional theological sense would be radically diminished. As I will try to show, although Kant devotes attention to the question of religion, his conception of God goes beyond the dogmatic framework of Christian doctrine, which influences his overall perception of the relationship between God and man. In particular, he questioned the theological assertion that God's grace is essential when considering the philosophical basis of moral behavior.

Spener is regarded as one of the most significant Protestant theologians of the modern era. However, comparing Kant's and Spener's thought is mostly beneficial for researching Kant's philosophy since the innovative nature of some of Kant's ideas – which do not necessarily seem revolutionary today – can be fully appreciated by comparing his thoughts with earlier traditions. Analyses of Kant's thought frequently occur in an isolation from its historical context: Kantian

philosophy is often interpreted as an autonomous system, with little consideration given to its position within the broader constellation of historical philosophical theories[1], the sources of Kant's inspiration, or the subtleties of his thought resulting from the historical evolution of the German language[2]. This approach may give researchers some autonomy in formulating their own interpretations, but ultimately leads to a kind of detachment of Kant's system from its historical context[3]. However, in the case of the topic of this paper, it is this context that is particularly important.

A comparison of Kant's and Spener's views on the means of achieving a better future reveals some interesting aspects. A thesis may be formulated that in the period of several decades between Spener's and Kant's activities, a specific notion of the future emerged, according to which humanity does not need supernatural assistance to achieve future goals. The assumption of independence from divine influence in shaping the future cancelled the limits of expectations rooted in traditional theology. This opened a new perspective on the enhancement of human morality, unrestrained by the doctrine of the original sin. Therefore, the secularization of eschatological expectations concerning the future, which may be observed in Kant's philosophy, may be treated as a systematic source of Enlightenment optimism concerning the future. Nevertheless, as will be demonstrated, the belief in humanity's capacity for self-determination in shaping its own destiny represents merely one of the factors that can influence the formation of an optimistic worldview. Another significant factor affecting the emergence of optimism is the conviction that – despite the lack of control over the future – the reality is progressing in an optimal direction independently of the influence of individuals[4]. This attitude can be defined as a form of trust in fate, or in other words, as a belief in an order of events that occurs spontaneously in an objectively correct manner.

Let us now undertake a detailed examination of Kant's and Spener's concepts to determine the extent to which their insights influence contemporary thought.

Spener and the Hope for a Better Future

Philipp Jacob Spener is most commonly known as the father of Pietism. Indeed, he made a significant contribution to the establishment of the entire pietist movement, as his ideas became the theoretical basis for the foundation of pietist doctrine. Spener's system is worthy of further consideration for one additional reason. Namely, Spener contributed to the establishment of a novel quality in theology, which also impacted early modern and Enlightenment philosophy. His

[1] If these theories are taken into account, they are usually analyzed to a much lesser extent than Kant's philosophy.

[2] The method I am referring to involves carefully and systematically reconstructing the thought systems of the authors, who shaped the intellectual landscape from which Kant's system emerged. As for the possible influence of Spener's thought on Kant, it is noticeable to some extent; however, Spener's theology is often reduced to a general understanding of Pietist doctrine (Cf. PASTERNACK, 2014, p. 169 –170). Any efforts to reconstruct this aspect are worthy of praise. However, it must be emphasized that each Pietist thinker held specific views that are rarely given due consideration and are often summarized too broadly. Therefore, considerations regarding such simplified concepts of pietistic theology do not provide a relevant picture of the complex theological movement we are dealing with when we speak of pietism.

[3] In the research paradigm, within which the presented analysis is conducted, Kant's system is considered historical philosophical theory developed over several decades without any claim to complete coherence, which is why a philosophical analysis of Kant's thought is not a final exegesis, but rather a reconstruction of Kant's general position in light of what can be historically established.

[4] This is a thesis that may not be irrefutable to researchers of Kant's thought. Dennis Vanden Auweele is a case in point. In his monograph Pessimism in Kant's Ethics and Rational Religion, he shows that Kant's views need not be interpreted as optimistic. (Cf. AUWEELE, 2018). However, my aim is to look at Kant's through the prism of the Enlightenment, which was characterized by a particular enthusiasm for the prospect of influencing reality.

theology was distinguished by an exceptional sophistication in addressing anthropological questions, which enabled him to elevate theological reflection to a level at which practical philosophy could readily engage with dogmatic issues. This process can be identified with the overcoming of the so-called metaphysical indeterminacy (CRISP, 2014, p. 81).

Spener's views on the future integrate several systematic elements. First, Spener held the conviction that the potential for individual moral advancement was a real possibility. The enhancement of human nature occurs in individual cases and is enabled by the event called "Wiedergeburt" – literally "rebirth" – which means "renewal". The concept of Renewal, also known as regeneration or restoration in theology, was not initially developed by Spener; it was already present in earlier theological theories for example in the thought of Martin Luther or in the Arminian theology. However, Spener proceeded to elucidate the central tenet of his system and subjected it to a meticulous analysis. In essence, for Spener, renewal signifies an instantaneous transition of an individual from the state of corruption caused by the original sin to an entirely new condition, which he designated as the state of a new human being. In his Pia Desideria Spener presented the famous sentence[5]:

> I will refrain from commenting on the other remarks made in the sermons. However, I consider this to be the most significant aspect, as the essence of Christianity lies in the inner or new human being, whose soul is characterized by faith and whose actions are guided by the fruits of this faith, which manifest as the fruits of life. It is, therefore, evident that the overarching objective of the sermons should be to achieve this outcome (SPENER, 2005, p. 162).

The transformation into a new human being was one of the most important elements of Spener's theology. The transition was posited to occur as a result of divine grace and to entail a complete reconfiguration of the cognitive and volitional structures of the human being. The character of this transformation is so all-encompassing that, according to Spener, the affected individuals literally become "new men". The consequences of original sin were partially negated, resulting in a partial restoration of both their condition and volition to their original state. As a result, the individuals who have undergone this transformation are able to comprehend the necessity of following the divine rules and develop the subjective desire to do so. Consequently, the range of decision options to which the will may be voluntarily directed is radically expanded.

In contrast to those who have not yet undergone the renewal process, the individuals who have been reborn are able to not only desire what is sinful but also to voluntarily strive for what is good. Nevertheless, they are not compelled to do so. They are able to choose freely between obedience and disobedience towards divine rules. In contrast to the reborn individuals, the volition of those who had not yet experienced renewal was consistently oriented towards sinful options due to the corruption of their nature by original sin. The doctrine of renewal was a systematic basis for Spener to develop the thesis about human freedom. According to Spener, the reborn human individuals can be acknowledged as free, since they can make their decisions completely indeterminately and independently of all external factors. It is precisely on this point that the high degree of innovation of Spener's system within the spectrum of theological theories of Lutheran provenance can be clearly seen: Luther's claim that man's free will is dead lost its validity in the face of Spener's interpretation of the doctrine of renewal.

[5] Self-translation. The original text is as follows: "Was ein und andere anmerckungen sonsten sind / die bey den Predigten zu beobachten / übergehe hier gern. Das vornehmste aber achte ich dieses zu seyn / weil ja unser gantzes Christenthum bestehet in dem innern oder neuen menschen / dessen Seele der Glaube und seine würckungen die früchten deß lebens sind: Daß dann die Predigten insgesampt dahin gerichtet solten warden" (SPENER, 2005, p. 162).

Consequently, Spener's conviction regarding the potential for individual moral advancement also encompassed the possibility of collective human progress. His views are most clearly exemplified by the title of one of his works, *Hoffnung besserer Zeiten*, which translates literally as *The Hope for the Better Times* (SPENER, 2001). It is not without reason that Spener uses terms such as "*Hoffnung*" ("hope") or "*Pia desideria*" ("pious wishes"), which suggest that success in improving the state of reality does not depend solely on man. Spener's principal objective was to elucidate the manner in which individuals might transcend their baser instincts in order to approximate a pious ideal to the greatest extent possible during their earthly existence. He held the conviction that the restoration of the dispositions of human nature corrupted by original sin in individual cases would enhance the general situation of all humanity. The collective moral progress of humankind would result in a state that is as similar as possible to that of the eternal Kingdom of God. In this regard, Spener's exhortation to believers to engage actively in the improvement of reality represents one of the most distinctive elements of his theology. The individuals had the real potential to influence the future.

However, Spener's vision of the future was not as straightforward as it may appear. Despite the assumption that individuals could influence the shape of the future world, Spener was convinced that the biblical vision of the approaching apocalypse could come true at any time. A number of passages in his works allow us to assume that he held a strong conviction that the era in which he lived may be the last. For example, he presented an intriguing interpretation of the Book of Revelation in his treatise *Muhammedismus in angelis Euphrataeis S. Johanni Apocal. IX. v. XIII. & seq. praemonstratus* (SPENER, 1664)[6]. It is surprising that Spener believed that individuals were in a position to influence the future, given his deep conviction about the possibly approaching apocalypse and the necessity of submitting to God's grace. It appears that there is a certain discrepancy in Spener's views on the future. On the one hand, he asserted that individuals could play an active role in improving the world. On the other hand, his theological views on the apocalypse implied that the fate of humanity depended largely on God.

What initially may appear to be a systematic contradiction in Spener's system of thought is, in fact, not an inconsistency at all. In order to ascertain the coherence of Spener's beliefs, it is essential to consider them from the perspective of contemporary modes of thought, which diverge considerably from the understanding of reality prevailing in the early modern and enlightenment epoch. A distinctive feature of Spener's perspective is its synthesis of the belief in the human capacity to shape future outcomes with the assertion that divine agency exerts a significant influence on these developments. This conviction in the capacity of humans to influence reality while acknowledging the role of a higher power led to the conclusion that, while humans should strive to shape the world, they are not the sole determinant of its form. The realization that despite their endeavors, they would not be the ones to determine the ultimate shape of the future remained an element present in Spener's thought but no longer a tenet of Kantian philosophy.

Kant and the Demythologization of Christian Anthropology

A direct comparison of the views of Spener and Kant reveals the emergence of a new element in philosophy over several decades: the idea of a possibility that the future may be solely deter-

[6] This is admittedly a very early text, but there is evidence that Spener already during his early activity represented views on freedom that are similar to those he put forward in his later writings. He presents such views for example in his *Dissertatio de Conformatione creaturae rationalis ad Creatorem* of 1653.

mined by humans. The idea was innovative in Kant's time because it offered an alternative to the previously dominant vision of the future, in which humanity had limited control over events because the future was largely determined by divine will. Spener's assumption of the possibility of moral progress can be said to have formed the basis of Kant's own thesis. Kant correctly identified that Spener, like himself, espoused the belief in the possibility of moral advancement. In *The Conflict of the Faculties*, Kant explicitly referenced Spener's concept of renewal.

> The problem (which the valiant Spener called out fervently to all ecclesiastical teachers) is this: the aim of religious instruction must be to make us other men and not merely better men (as if we were already good but only negligent about the degree of our goodness). This thesis was thrown in the path of the orthodox (a not inappropriate name), who hold that the way to become pleasing to God consists in believing pure revealed doctrine and observing the practices prescribed by the church (prayer, churchgoing, and the sacraments) to which they add the requirement of honorable conduct (mixed, admittedly, with transgressions, but these can always be made good by faith and the rites prescribed). The problem, therefore, has a solid basis in reason. (SF, 7:54)

As can be observed, Kant's stance on this matter is not entirely uncritical. His criticism was fundamentally directed against Spener's claim that the character of an individual's transformation from the state of the corrupted nature to the state of a new human being is supernatural or must be recognized as a "miracle". The very element in Spener's system that triggered Kant's criticism was thus the claim about human beings' dependence on divine support in the face of their moral progress. Kant's position was that individuals are capable of attaining moral perfection independently, given that the capacity to do so is inherent in their rationality and can be fully explained within the framework of the normative nature of moral laws.

Kant says the following:

> But the solution turns out to be completely mystical, as one might expect from supernaturalism in principles of religion; for, according to it, the original, incorruptible moral predisposition in man's nature, though supersensible, is still to be called flesh because its effect is not supernatural as well; only if spirit (God) were the direct cause of man's improvement would this effect be a supernatural one. So man, being by nature dead in sin, cannot hope to improve by his own powers, not even by his moral predisposition. (SF, 7:54)

When viewed through the prism of early modern soteriological theories, it may be seen that Kant believed that humanity possessed some capacity to conquer evil without reliance on supernatural assistance. The theological narrative that people can aspire to goodness, but always require divine intervention to really achieve it, was problematic for Kant, since it implied an assumption of an irremovable passivity of humanity in its quest to improve its moral condition

This is a thesis that is certainly worth pausing over, as it may raise some objections.[7] Firstly, Kant did not claim directly that people's capacity for moral goodness implies that they will certainly act morally. Similarly, it does not follow that the propensity to do evil can suddenly be eradicated from human nature. Furthermore, Kant drew attention to the collective dimension of human activity, which conditions the emergence of space for morally good actions: Although each individual independently engages in moral reflection and makes individual decisions, Kant is also aware that the collective aspect plays a certain role in the process of moral progress.

[7] As Pasternack points out, this is a question in Kant's philosophy that is difficult to resolve unequivocally. Pasternack says: "(…) in Religion this particular need for Divine assistance is removed. With the introduction of the Change of Heart, we do not need to be forgiven for falling short of a standard that cannot possibly be realized, for it is a standard that is actually within our reach. Nevertheless, this does not preclude some other necessary role for Divine aid. But unfortunately, the text is notoriously unclear on this point, with various passages conveying different views as to whether it is still in some way necessary, or whether we can accomplish the Change of Heart fully on our own" (PASTERNACK, 2014, p. 144).

Nevertheless, Kant argues that it is unjustified to expect individuals to act in the right way if they are not capable of doing so unassisted – in this point his views are completely contrary to what theological theories have claimed, regardless of the assumed degree of human capabilities in a soteriological perspective in those theories. Moreover, for Kant the belief that evil can be possibly overcome justifies the principle found in Christian moral teaching: the fact that God requires humanity to be good implies that humanity must be capable of this. It is groundless to require of someone something that is beyond their capacity.

In *Religion within the Boundaries of Mere Reason,* Kant says the following:

> For, in spite of that fall, the command that we ought to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it, even if what we can do is of itself insufficient and, by virtue of it, we only make ourselves receptive to a higher assistance inscrutable to us (RGV, 6:45).

The revolutionary nature of Kant's statement can easily be overlooked if the exact context in which he presents his statement is not taken into account. It is significant, that Kant does not utter these words only in the general context of his reflections on religion, but that this is one of the few places in his writings where he refers to a specific theological concept. Since Kant formulated his views in the context of Lutheran theological tradition, particularly Pietist theology, he was responding to the long-standing theological heritage of viewing human autonomy in doing good as being extremely limited by the necessity of God's support. It is precisely the context that is crucial to fully appreciate the uniqueness of Kant's thesis, and it should not be overlooked when analyzing the quoted passage.[8]

It is not without significance that Kant utters these words in relation to the problem of theodicy, directly referring to the legacy of Gottfried Wilhelm Leibniz. In doing so, Kant touches upon the very core of the problem of world's dependence on God's will. However, the view presented in the quoted passage changes radically over the next thirty years (Cf. KROUGLOV, 2018). Kant distanced himself from a way of perceiving reality, in which man is treated merely as an element of created reality, and eventually attributed to man some power to shape that reality. It is also evident that, as time passed, Kant's approach to articulating the question of religious matters underwent a shift and the image Kant presents in his later reflections is already different.[9]

In order to understand the idea of demythologizing Christianity in Kant's thought, one must pay attention to one more key aspect. Kant perceives the idea of God in a completely innovative way, treating it as an inherent element of a philosophical system (Cf. THEIS, 1994). The concept of God is therefore not the dogmatically determined foundation on which the system is built, but is itself part of that system (see THEIS, 1994, p. 323).[10] This may be one of the reasons,

[8] In this context, interesting perspectives are provided by recent monographs entitled: *Religionsphilosophie nach Kant: Im Angesicht des Bösen* edited by M. Kühnlein and *Kants Theorie des Bösen im Kontext* edited by G. Sans and T. Hanke (Cf. KÜHNLEIN, 2023; HANKE, 2024).

[9] Not only did the position of man come to the fore, but also the very concept of God evolved: It developed from the idea of a real entity – which Kant discussed treating Wolff'' metaphysics and Newton's physics as his point of reference – to the regulative idea (Cf. THEIS, 1994, p. 33– 34; p. 325).

[10] At this point, it is worth quoting Theis directly: "Unsere gesamte Rekonstruktion der Entwicklung von Kants theologischem Denken orientierte sich am Gedanken einer Konsistenz bzw. Harmonie zwischen verschiedenen Diskursebenen innerhalb eines philosophischen Gesamtprojektes. Dieser Grundgedanke war zunächst als ein Postulat verstanden oder gedeutet worden, nämlich derart, daß ein philosophischer Diskurs seiner Idee nach konsistent sein soll. Unsere These lautete, daß bei Kant der theologische Diskurs ein Teil eines philosophischen Gesamtprojektes ist und daß er demzufolge nicht unabhängig von diesem funktioniert"

why Kant's understanding of God and of matters of faith in general may sometimes appear to be inconsistent with Christian dogma – however, Christian dogma did not have the same systematic importance for Kant as it did for earlier thinkers. It is not difficult to see that Kant's view of man's position in the world, but also his way of expressing himself on the truths of the Christian faith, begins to deviate significantly from the paradigm dominant in Christianity. It can, therefore, be reasonably assumed that Kant presented already views one could consider secularized – even if religious matters still remained an important subject of his reflections. The use of the term "secularized" seems appropriate in this particular sense, as Kant's perspective on issues of faith is not dictated by any ecclesiastical authorities; rather, it emanates from Kant's aspiration to preserve the coherence of his own philosophic system.[11]

Despite the differences between Kant's and Spener's views, it is difficult to deny that both thinkers were convinced of the real possibility of individuals' moral progress. It is, therefore, pertinent to inquire about the implications of Kant's assumption that the future is solely within the control of humanity and should not be treated as a subject of divine influence. In other words, what were the consequences of Kant's secularized view on the possibility of future improvement?

Kant's Secularized View and its Consequences

The initial consequence is a particularly straightforward implication. Kant's concept of moral progress for individuals excludes the assumption about the remaining traces of the original sin in human nature, which was a common viewpoint among theologians. For Spener, the concept of renewal does not imply the complete eradication of the corruption of human nature resulting from original sin; rather, it entails a state of equilibrium. Unlike Spener, Kant did not believe in the concept of original sin in the traditional theological sense: his understanding of moral progress was that it involved the ultimate elimination of evil in human nature, which occurs only as a result of an individual's decision to consistently respect the moral law. Consequently, Kant's perspective is more optimistic than Spener's, who considered it implausible to eradicate the vestiges of original sin in earthly life (SZYRWIŃSKA-HÖRIG, 2024, p. 44–46).[12]

The second consequence of the secularized view on the future is that the considerations re-

(THEIS, 1994, p. 323). The distinction between the theological and secular perceptions of religion should however not be confused with Kant's own distinction between historical faith and a purely rational religious system.

[11] At first glance, Kant's views seem to be completely immersed in a religious context.: Kant speaks of the existence of God and the significance of establishing the Church. One can even point to interpretations that equate Kant's concept of the church with his idea of kingdom of ends (see ex. PALMQUIST, 1994, p. 426). However, in a historical context, his way of expressing himself on theological figures is a milestone. It suffices to look at his claims through the prism of the philosophical reflections of German thinkers – not to mention theological theories – in the 17th century to see that Kant represents a qualitatively different view of religion than his predecessors, ex. Christian Wolff and Christian August Crusius. Kant decisively rejects a certain paradigm, still derived from scholastic philosophy, of speaking about religious categories in a manner consistent with doctrine, and begins to speak about them in a completely independent manner. As previously mentioned, religious concepts are an element of the system for Kant, not a doctrinally defined external reference point. The consideration (or lack thereof) of this distinction by contemporary commentators on Kant influences their interpretation of his thought. For example, it is worth comparing the different views on Kant's ethics presented by J. B. Schneewind and J. Hare. Although Hare accuses Schneewind of misinterpretation, Schneewind analyses Kant's thought from a different perspective. While Hare makes operational comparisons between Kant's thought and that of earlier philosophers, showing how their theories relate to each other, Schneewind considers both changes in Kant's thought over time and changes in the paradigm that occurred in his philosophy. (Cf. HARE, 2000, p. 463; SCHNEEWIND, 1997, p. 512-513).

[12] It should not be forgotten that Spener, although he made significant modifications to Luther's theology, professed the characteristic Lutheran belief in the inability of humans to improve their condition without God's help.

garding the ideal future state in Kant's philosophy have shifted from the theological dimension to the political, which may be observed for example in his theory of peacemaking (Cf. PALMQUIST, 1994, p. 422). If one considers that the earlier theories of peace were invariably inextricably linked with the examination of the legitimacy of certain religious beliefs and the debate concerning their dissemination, the innovative nature of Kant's project of perpetual peace becomes evident. Although Kant speculated about the enhancement of individual moral condition, his vision of the future does not contain any kind of practical advice on what exactly individuals should or could do in their daily lives in order to experience their moral metamorphosis. His remarks on the positive transformation of the basis on which individuals accept maxims are very abstract in nature and cannot be considered as pragmatic advice (RGV, 6:44–53). Instead, Kant developed the concept of perpetual peace, which had an entirely different character than Spener's instructions concerning the individual strategies to contribute to the enhancement of the future world. Therefore, it is legitimate to argue that the overarching Kantian project of future improvement was primarily concerned with the problem of governing social and political reality, with relatively little attention paid to the pragmatic dimension of pursuing individual moral progress. At this point, it becomes evident that there is a significant discrepancy between Spener's notion of the strategies for the enhancement of the future, which was predicated on the individual moral progress contributing to the betterment of life for the entire human race, and Kant's assertion that the formulation of general rules on the political level represents the optimal means of achieving the ideal future state in which the entire human race could flourish. Although Kant does not explicitly state this, there are a number of reasons to believe that his project for perpetual peace was influenced by his secularized anthropological notion. One particularly notable assumption is Kant's treatise entitled *Toward Perpetual Peace*, about which one may legitimately assume that it is rooted in his secularized view of the possibility of individual moral progress excluding the support of a higher power. In essence, Kant was convinced of the ultimate sustainability of peace. It is crucial to emphasize that Kant's concept of peace was not merely a temporary cessation of military hostilities; rather, it encompassed the absence of any potential for conflict to arise. Kant spoke in a tone that suggests that it is possible to achieve a state of peace that is permanent in nature. Interestingly, he suggests that one of the conditions for bringing about peace is the abolition of armies. This refers to the so-called third article of the Preliminaries, which begins with the words:

> Standing armies (*miles perpetuus*) shall gradually be abolished entirely for they continually threaten other states with war by their willingness to appear equipped for it at all times. They prompt other states to outclass each other in the number of those armed for battle, a number that knows no limits. And since the costs associated with maintaining peace will in this way become more oppressive than a brief war, these armies themselves become the cause of offensive wars, carried out in order to diminish this burden (VAZeF, 08:345).

This is one of the numerous conditions for perpetual peace presented by Kant, which is, however, of a unique nature. It suggests that the mere removal of the potential for military operations could indeed contribute to the extinction of conflicts. While the abolition of armies is a necessary first step on the path to achieving perpetual peace, Kant's premise is based on the thesis that the source of conflict can be extinguished. Such an assumption is systematically incompatible with the classical theological doctrine of original sin. The classical dogma of original sin posits that the fallen nature of humankind is a source of potential conflict that cannot be rectified. This can be observed in the theology of Spener, who claimed that even the nature of the reborn individuals was, to some extent, influenced by the original sin. As long as the effects of the original

sin – commonly referred to as "maliciousness" – remain within the nature of human beings, the potential for conflict will persist. From this perspective, even the complete abolition of all military forces, whose existence could be perceived as a threat to others, does not necessarily eliminate the underlying causes of this potential hostility. An interesting point emerges here: if Kant did not deny the classical theological view on human nature, which implies the impossibility of liberating humanity from the consequences of original sin, he would not be able to assume the possibility of achieving sustainable peace as the potential source of the possible conflicts among people would never be eliminated.

It is crucial to emphasize that this is an aspect of Kant's thought that is most clearly evident in the analysis of Kant's philosophy in relation to the historical context of its emergence. The fact that Kant even reflects on the potential possibility of permanently extinguishing conflicts appears to be unique in comparison with earlier theories of peace, which focused on introducing harmony between warring parties but presented a pessimistic view on the elimination of the foundations of conflict inherent in human nature, whether original sin or religious differences.

Spener's and Kant's Notions of Optimism

The differences between Kant's and Spener's versions of optimism become apparent when one looks at their views on optimal strategies for preparing for the future. Spener urged believers to embrace conversion and atonement, which he believed would signal to God that they were prepared for renewal. With divine assistance, they could undergo a moral transformation, and as the number of those who had been reborn increased, the entire human race would begin to resemble the Kingdom of God. This condition could potentially be interrupted by the apocalypse, about which Spener was convinced it was approaching. Nevertheless, even if God decided to induce the apocalypse, the efforts of humankind would not have been in vain. The pursuit of renewal and the endeavor to align the terrestrial realm with the Kingdom of God represented the optimal and sole means by which humankind could prepare for the future. Spener's views were thus a synthesis of optimism and belief about the possibility of progress on the one hand and humility regarding the independent decisions of God on the other.

In Kant's philosophy, the theologically determined idea of necessary subordinance towards God has been superseded by a strong emphasis on pragmatism. Kant believed that humanity has great autonomy and self-governance over its reality. While he did not exclude the possibility of divine intervention, he did not consider it as a factor that should be taken into account in the rational analysis of the current state of affairs or the design of the future. This is because, as has already been emphasized, Kant assumes that if it is reasonable to expect humanity to become morally better, then this must be achievable for humanity (RGV, 6:45). Even if divine intervention was possible, Kant maintained that humanity should be responsible for the future as if everything depended solely on its own decisions (Cf. ex. RGV, 6:28–32; 40–41).

A further notable divergence between Spener's and Kant's perspectives on the enhancement of the future pertains to the potential for radical transformations in human nature. Both authors were convinced of the capacity of humans to undergo positive transformation. Spener advanced a theological concept of spiritual renewal, which constituted the focal point of his theological thought. Similarly, Kant believed that humanity should be acknowledged as capable of struggling against the evil inherent in its nature without the assistance of supernatural means. The

divergence between the two authors' perspectives on the permanence of this overcoming of evil was a significant point of contention. Spener was faithful to the Christian doctrine and believed that the effects of original sin on human nature could never be completely eradicated. Instead, he argued that they could only be mitigated through divine grace. Kant, however, was much more radical in his views, proposing that the individual could overcome the evil in their nature by making a conscious and voluntary decision to act consistently according to morally good maxims (RGV, 6:44–53). Although Kant does not rule out the possibility of God's support in doing good, he takes a revolutionary step by creating space for discourse that does not take this help into account.

The differences between the versions of optimism that characterize the thoughts of both authors are becoming apparent. Spener's optimism is distinguished by a certain degree of moderation with regard to both the capacity of individuals to act independently in shaping the future and the potential for radical moral improvement. Although such an improvement is a possibility and should be striven for, people remain highly dependent on divine grace. Similarly, all efforts to shape the future for the better should, according to Spener, proceed with the understanding that the ultimate outcome of the future is not under the control of human beings but is instead a matter of divine determination. These beliefs imbued Spener's perspectives with a notable moderation, a quality that stands in stark contrast to the more radical stance observed in Kant. Kant allowed for the possibility that humanity may be solely responsible for shaping its own destiny – people's decisions should be made with the awareness that they influence the shape of reality, rather than merely cooperating with divine intervention. Similarly, he held the view that the individuals were capable of overcoming the evil within themselves.[13]

In Kant's thought, a significant development was made: the assumption dominating over centuries that humankind is dependent on God was treated as unobvious. In Kant's philosophy, the concept of future thinking underwent transformation, evolving from a contemplative exercise to a pragmatic design for a better world that was devoid of its traditional sacred dimension. The vision of the apocalypse has been superseded by a vision of a better tomorrow. Similarly, the millenarian expectation of the *Second Coming* has become an expectation of the advent of an era of a better life. The process of humanity's emancipation from a higher power, symbolized by God, can be viewed as an exemplification of the Enlightenment ideology, which underscored the intrinsic potential of human nature. Consequently, the secularization of future perspectives can be regarded as an indicator of progress in human thought and, as Kant himself asserted, liberation from "self-incurred immaturity" (WA, 8: 35).[14]

In order to identify the conclusions that can be drawn from the analysis of historical theories which could be of benefit in the modern world, two key issues require consideration. The initial issue for consideration is that of modern optimism, or more specifically, the question of whe-

[13] This becomes clear when we examine Kant's erroneous criticism of Spener's idea more closely. A thorough analysis of this issue is required, but briefly, it can be presented as follows: Kant believed that, according to Spener, humanity would overcome evil through renewal ("Wiedergeburt"). However, he criticized the fact that, according to Spener, Renewal could only take place through divine intervention. Therefore, Kant believed that improving the moral condition must depend on human beings themselves. However, he did not take into account that, for Spener, Renewal did not signify the total eradication of evil. By misunderstanding Spener's idea, Kant revealed his own views on the possibility of overcoming evil in human nature.

[14] When speaking about progress, it is not my intention to suggest that secularized views are inherently superior to those shaped by religious convictions. Rather, I am simply emphasizing the changes that have occurred within the field of philosophy, without making any evaluative judgements.

ther contemporary society shares the view held by Kant and Spener that the future of humanity will be better and that some enhancement of human nature may take place. The second issue pertains to the question of the sense of independence from external factors that accompanies individuals in their daily lives. It is about the extent to which people feel able to plan and shape their own future. By examining these two issues, we can attempt to ascertain whether, in relation to them, we can speak of a modern inheritance of the way of thinking that was present in the thought of Spener and Kant. We may, therefore, consider whether the contemporary perception of the world corresponds more closely to the thought of Spener or Kant.

The Contemporary Form of Optimism

Historical analysis provides a valuable opportunity to challenge past theoretical perspectives with contemporary thinking. All the more so when reflecting on optimism since contemporaries – just as before – look to the future with certain expectations, hopes and fears. However, it is difficult to give a clear answer to the question of the determining factor of their worldview. Therefore, in order to gain an insight into the nature of modern optimism[15], let us take a look at the question of modern optimism and compare it with the optimism that characterized the thinking of Kant and Spener.

When trying to answer whether contemporary thinking about the future is characterized by optimism, a certain ambivalence emerges at first glance. On the one hand, we like to think of ourselves as the successors of the Enlightenment. Its demands for rationality, a universal system of knowledge and communication as a means of exchanging information are very close to our own. Like the Enlightenment, we seek to combat superstition and irrational prejudice. A common approach to religiosity is a compromise between faith and rationality. All these features seem to indicate that, as in the Enlightenment, we should believe in progress and in the possibility of shaping reality for the better in the modern age. Plenty of evidence suggests that we have every reason to be optimistic about the future. Particularly significant here is the recognition of the developments in technology and science that are opening up perspectives for dealing effectively with problems that have posed real challenges in the past. We are aware that science and technology are continuing to develop, giving hope that the methods of dealing with problems will become increasingly effective in the future.

At the same time, however, an optimistic view of the future is not definite. There is no denying that the current sense of being able to have a positive impact on the future is accompanied by a certain fear of the dangers that the future may bring. At this point, it is sufficient to mention three such threats that are predominant. The first is climate change, the second is the prospect of epidemics, and the third is the potential for the outbreak of war. It is worth noting that none of these is in itself new: wars and epidemics have occurred in the past, and the vision of ecological catastrophe due to climate change and pollution has been developing over decades. Nevertheless, it is evident that the past few years have served as a stark reminder to individuals across the globe that these perils are not merely hypothetical but rather have the tangible capacity to impact us. It becomes evident that the discussion about the aforementioned threats does not con-

[15] I use the term "modern optimism" to refer to a number of observations concerning the modern perception of reality which have been made in recent years, ex. in Josef Römelt's article "Zwischen medizinischem Optimismus, sozialer Distanzierung und religiöser Lebenshilfe". Earlier, Oskar Pfister also provided interesting comments on optimism and the sense of having an impact on reality in his article *Die Rolle des Unbewussten im philosophischen Denken*.

cern merely potential possibilities; rather, it encompasses options that could materialize in the near future and whose effects will directly affect the contemporary, as well as future generations. As a result, a heightened sense of insecurity is evident. Furthermore, it is notable that these threats manifest themselves in diverse dimensions of reality. For instance, while an epidemic is a purely biological phenomenon, the outbreak of war is already an event that can be considered from a political and ethical perspective. This illustrates that the dangers we currently perceive have their roots in both the natural world and human nature.

It can, therefore, be argued that the optimism characterizing contemporary thinking stems from the belief that individuals have the ability to modify their situation according to their own preferences. The belief that people can actively change their future makes them optimistic that this future will be better. However, this is not an unconditional form of optimism. The dominant mood of our era is namely that of decline, which serves as a counterweight to the sense of optimism. This observation justifies the thesis that current pessimism has its genesis in a pervasive sense of the lack of control over various dimensions of reality that may be perceived as threatening. Despite technological advancement, it is not feasible to circumvent the risks associated with biological processes. Similarly, it seems that the moral and social values espoused are not sufficiently universal to avert potential political and military conflicts.

Nevertheless, an examination of the ideas put forth by Spener and Kant reveals that both thinkers also did not assume the complete independence of humanity in planning the future. In Spener's system, God was the primary determinant of the future world. Spener's system advocated concern for the future yet simultaneously asserted that the ultimate form of reality was contingent upon divine intervention. Kant's thought diverges from this perspective. He posited the autonomy of humans from supernatural influences, ascribing to them a markedly elevated degree of autonomy in the shaping of the future in comparison to Spener's perspective. Kant held the view that individuals are capable of independently influencing their moral development and even overcoming the morally negative aspects of their nature. This conviction also determined his views on progress in interpersonal reality. Kant postulated using the potential for the shaping of the social and political dimensions of life, as well as the establishment of a permanent state of peace on Earth. The significance of this move is difficult to overstate; one might even suggest that Kant liberated eschatology from its temporal religious context. Nevertheless, even Kant assumed that the future of humanity is not solely within the power of humankind to determine since progress does not occur randomly but takes place within some concrete boundaries. In Kant's writings, there are passages that posit the idea that nature and history, understood as a development of reason, is the driving force behind human actions, carrying out its plan, for example:

> All of creature's natural predispositions are destined eventually to develop fully and in accordance with their purpose (IaG, 8:18).
> In the human being [...] those natural predispositions aimed at the use of its reason are to be developed in full only in the species, but not in the individual (IaG, 8:18).
> And thus is the outcome of an attempt to write, through philosophy, the most ancient part of human history: satisfaction with providence and the course of human events as a whole, a course which does not progress, beginning with good, toward evil, but rather develops gradually from worse to better. Everyone is called upon by nature itself to contribute, to the best of his ability, his part of this progress. (MAM, 8:123).

Kant was persuaded that human experience was an integral aspect of a universal historical process. This prompted him to ask the question of the meaning that accompanies history. On

an initial observation, the history of humankind is defined by a multitude of violent and immoral occurrences, including armed conflicts. Nevertheless, Kant maintained that it is worthwhile considering the possibility of an overarching objective towards which history is progressing. Kant thus concluded that in the life of individuals, as well as in the life of humanity as a whole, reason comes to the fore and constitutes the ultimate goal towards which everything strives. It is precisely the thesis of the development of reason and the strengthening of the position of reason in the world that can be taken as the teleological context for Kant's considerations. Therefore, according to Kant, despite humanity's control of reality, all human activity is part of a certain purposeful order. It can be argued that Kant's belief in the purposefulness of nature and history acts as a moderating factor of human freedom. Nevertheless, Kantian teleology allows for a considerable scope for human spontaneity. This represents an expression of the Enlightenment belief that progress is not accidental but rational and, therefore, a positive phenomenon.

Optimism Between Control and Trust

It is interesting to note that, looking at the views of people living in the present, it can be concluded that it is the sense of the loss of control over reality and the presence of uncontrollable factors that prevent the development of an optimistic attitude[16]. As has been demonstrated, epidemics, wars, and the dangers of a worsening climate are examples of factors that present a problem for an optimistic approach to life. All these phenomena are uncontrollable and cannot be considered with certainty to be avoidable. In order to comprehend the genesis of optimism, it is however important to make the distinction between exerting control over reality and the conviction that, despite the absence or strong limitation of control, events are progressing in the objectively most favorable manner. This disposition can be designated as trust in fate.

The differentiation between control and trust appears to be crucial in the context of investigating the source of optimistic beliefs. All indicates that it is not only a sense of control over reality that is the source of optimism but rather a kind of belief that events will unfold in a favorable way whether or not they correspond to the plans of individuals. It is important to note that the two factors of the sense of control over reality and belief in fate are not mutually exclusive; they do not necessarily condition each other. It is evident that the sense of control may coexist with trust in fate; however, this is not a prerequisite. Furthermore, there is some indication that belief is a more important factor than a sense of control in the formation of an optimistic attitude. The sense of control can be a contributing factor to optimism; however, it is trust in fate that is a necessary condition for its formation. The mere presence of the conviction that the world is heading in the right direction, even if only in spite of human influences, is sufficient for an optimistic view of the future. Nevertheless, the sense of control over reality devoid of faith in fate can invariably be accompanied by a pessimistic apprehension that this control may be misused, resulting in regression.

These theses are confirmed by the observations of modern researchers. Modern theories also deal with the theme of belief that the future will bring the expected positive events, but they do not speak of "trust in fate", but rather of "hope". It can be seen, however, that modern authors identify precisely with the concept of "hope" the potential of the individual to develop the conviction that the future with a certain probability will bring the expected positive events.

[16] Josef Römelt, for example, pointed out that an optimistic attitude could be at risk (RÖMELT, 2022, p. 33).

However, in the contemporary discourses the very concept of "hope" has different meanings. For example, Adrienne Martin distinguishes between hope understood "as a combination of the desire for an outcome and the belief that the outcome is possible but not certain" and the second type of hope, which she describes as "hope against hope". The latter type she describes as "hoping for an outcome that one highly values but believes is extremely unlikely" (MARTIN, 2014, p. 5). The important point is that both types of hope assume the impossibility of the individual to completely influence the shape of future events, while at the same time assuming that the individual believes that the future can bring what he or she expects, however improbable it may be. It is precisely the inability to control future events that is the condition for the subject to develop hope. Philip Pettit, in turn, emphasized the relationship between hope and the awareness of one's own limited possibilities in influencing future events:

> To hope that something is the case or that you can make it the case, then, is to form an overall outlook akin to that which would be appropriate in the event of the hoped-for scenario's being a firm or good prospect. Where the prospect is manifestly beyond your control, it is to sustain a more or less sanguine set of attitudes and to act on other fronts in the way that such attitudes would prompt. And where the prospect is within your sphere of influence – however improbable it is that the influence will be effective – it is to act as if there were also a good chance of making it come out as you wish. It is to embrace an assumption that gives you heart and life and energy. It is to embrace this assumption, furthermore, even if you happen to wax quite optimistic (PETTIT, 2004, p. 158).

Another important feature of hope emphasized in modern theories also corresponds to the early modern and Enlightenment notion of faith in fate. This is the belief that the future will be better, even if future events do not directly contribute to the personal satisfaction or happiness of the individual. Hope refers to the expectation of positive change in general, rather than the expectation that an individual will be fortunate. This aspect was emphasized by Adam Kadlac, who contrasted the belief in a better future with the individual's expectation of their own happiness. It is the belief in a better future, not identified with the individual's desire to achieve happiness, that is said to create the potential for the individual to develop an attitude of hope. Kadlac states: "[...] I want to suggest that the hope for a good future is better suited than the hope for happiness to constitute the general trait of hopefulness" (KADLAC, 2017, p. 213).

Despite the similarities in understanding individual expectations of a positive future, a specific difference between modern and Enlightenment types of such optimism needs to be noted. This is evident when we look at the terminology used by contemporary authors. As can be seen in the quotations, the central term for the expectation of a better future is "hope". In the article presented here, we did not use the term "hope" but instead referred to "trust". This is an obvious difference, and there is a good reason for it. This change in terminology is due to the fact that the contemporary considerations are carried out within a context which does not directly refer to the religious sphere. Therefore, the factor of destiny in the eschatological sense does not appear in it. By emphasizing this difference in terminology, it is easier to see that historical expectations of a better future have shown a stronger attachment to the idea that a better future is guaranteed by some external factor – be it God, the natural order of things or the development of reason. Therefore, in the historical context it is much more appropriate to speak of "trust", because trust refers to an external factor which is a guarantee of the eventual fulfilment of expectations. The use of the term "hope" does not presuppose any such factor. Thus, "hope" refers to a situation in which someone would very much like a certain scenario to come to pass but cannot say that they trust it to happen because they believe in the operation of certain laws. "Hope" is based on

probability, while "trust" is based on confidence in a certain higher order.

It is interesting to note that, as observed in the introduction, there is no empirical evidence to support the belief that the future will be inherently better than the present. Even the scientific evidence of human progress does not provide sufficient grounds for an unquestioning belief in a better future. Therefore, such a belief must have some other basis. In light of the aforementioned observations, it can be concluded that the foundation of an optimistic worldview is not contingent on a reality check or scientific data. Instead, it is shaped by a belief in the potential for a better future, which we defined as trust in fate. The question thus arises as to the source of this trust. On the basis of the analysis presented, it can be concluded that two distinct sources can be identified in early modern and Enlightenment thought. These are the belief in divine providence and the belief in the historical progress of reason.

Concluding Remarks

The aim of this article has been to identify the source of optimism in people's thinking about the future. It can be seen that this source is to be identified with a worldview that is based on the belief that the future will have the most optimal shape. What appears special, however, is that the belief that the future will be better has a specific basis. It is not an empirical basis, nor is it simply a hope that is combined with a wish for something that may be unattainable. It is much more a belief in a certain factor that ensures that things happen independently of individuals in the most optimal way. This is an interesting observation in the context of one fact. Namely, as we have noted, the time in which Kant and Spener formulated their visions of the future was a period in which there was a kind of liberation of human thinking about the future from religious eschatological tendencies. Kant can be regarded as one of the first thinkers, who attributed such a high degree of autonomy to humanity, that it could be assumed that the future lies in the hands of human beings and not only of God. This was, however, not his arbitrary interpretation of human condition, but rather the result of breaking away from viewing religious issues from a dogmatic perspective and instead placing them on an equal position with other elements of his philosophical system. It was precisely this consideration of God as an integral part of the philosophical system, rather than as an external point of reference against which the system was created, that constituted the most important step in laying the foundations for optimism in Kant's version.

This observation is noteworthy in that it situates the optimistic worldview on a par with the categories of religious thought. This observation provides insight into the defining characteristics of the Age of Enlightenment. If an optimistic attitude to the future was one of the most characteristic features of the Enlightenment epoch, and if this optimism was based on a belief in a better future that could not be substantiated by any evidence, then it can be concluded that Enlightenment optimism was more ideological than rationally justified. This conclusion may not appear satisfactory in light of the assumption that the Age of Enlightenment was a time in which a rational approach to reality played a special role.

Nevertheless, this fact sheds light on a significant aspect of human nature. It appears that perspectives on the future can evolve independently of the logical indications of the potential trajectory of future events. This phenomenon is worthy of further consideration, particularly in the contemporary era, where the confrontation with cultural and technological changes can evoke

sentiments of both optimism and anxiety. These sentiments may potentially motivate concrete actions that could have profound and sustainable effects for future generations.

Bibliographic References

AUWEELE, D. V. 2018. *Pessimism in Kant's ethics and rational religion*. Lanham: Lexington Books.

CRISP, O. D. 2014. *Deviant Calvinism:* broadening reformed theology. Minneapolis: Fortress Press.

HANKE, T.; SANS, G. 2024. *Kants Theorie des Bösen im Kontext*. Hamburg: Meiner Verlag.
HARE. J. 2000. Kant on recognizing our duties as God's commands. *Faith and Philosophy*, [*s.l*] v. 17, n. 4, p. 459 –478.

KADLAC, A. 2017. Hopes and Hopefulness. *American Philosophical Quarterly*, Minnesota, v. 54, n. 3, p. 209–221, jul 2017.

KANT, I. 2006. An answer to the question: what is enlightenment?. In: COLCLASURE, D.L. (ed.). *Toward perpetual peace and other writings on politics, peace, and history*. New Heaven and London: Yale University Press.

KANT, I. 2014. An attempt at some reflections on optimism. In: WALFORD, D. (ed.). *Theoretical philosophy, 1755-1770*. Cambridge: Cambridge University Press.

KANT, I. 2006. Conjectural beginning of human history. In: COLCLASURE, D.L. (ed.). *Toward perpetual peace and other writings on politics, peace, and history*. New Heaven and London: Yale University Press.

KANT, I. 2006. Idea for a universal history from a cosmopolitan perspective. In: COLCLASURE, D. L. (ed.). *Toward perpetual peace and other writings on politics, peace, and history*. New Heaven and London: Yale University Press.

KANT, I. 1992. *The conflict of the faculties*. London: Nebraska University Press.

KANT, I. 2006. Toward perpetual peace: a philosophical sketch. In: COLCLASURE, D. L. (ed.). *Toward perpetual peace and other writings on politics, peace, and history*. New Heaven and London: Yale University Press.

KANT, I. 2018. *Religion within the boundaries of mere reason*. Cambridge: Cambridge University Press.

KROUGLOV, A. N. 2018. Kant and the crusians in the debate on optimism. *Kantian Journal*, Kaliningrad, v. 37, n. 2, p. 7–31.

KUPŚ, T. 2008. *Filozofia religii Immanuela Kanta*. Toruń: Wydawnictwo Naukowe UMK.

KÜHNLEIN, M. 2023. *Religionsphilosophie nach Kant:* im Angesicht des Bösen. Berlin: J.B. Metzler.

MARTIN, A. M. 2014. *How we hope*: a moral psychology. Princeton and Oxford: Princeton University Press.

PALMQUIST, S. 1994. "The kingdom of God is at hand!" (did Kant really say that?). *History of Philosophy Quarterly*, Illinois, v. 11, n. 4, p. 421–437, oct 1994.

PASTERNACK, L.R. 2014. *Routledge philosophy guidebook to Kant on religion within the boundaries of pure reason*. London and New York: Routledge.

PETTIT, P. 2004. Hope and its place in mind. *The Annals of the American Academy of Political and Social Science*, [*s.l*], v. 592, p. 152–165.

PFISTER, O. 1949. Die Rolle des Unbewussten im philosophischen Denken. *Dialectica*, Geneva, v. 3, n. 4, p. 254–271, dez 1949.

RÖMELT, J. 2022. Zwischen medizinischem Optimismus, sozialer Distanzierung und religiöser Lebenshilfe. In: BAHNE, T.; RÖMELT, J. (eds.). *Lebenswert in Verantwortung*: ethische Herausforderungen in der Corona-Pandemie. Würzburg: Echter Verlag.

SAVAGE, D. 1991. Kant's rejection of divine revelation and his theory of radical evil. In: ROSSI, P. J.; WREEN, M. (eds.). *Kant's philosophy of religion reconsidered*. Indianapolis: Indiana University Press.

SCHNEEWIND, J. B. 1997. *The invention of autonomy:* a history of modern moral philosophy. Cambridge: Cambridge University Press.

SPENER, P. J. 2001. *Hoffnung besserer Zeiten:* Erwartungshorizonte der Christenheit. Drei Schriften Philipp Jakob Speners aus den Jahren 1693/94. Hildesheim: Georg Olms Verlag.

SPENER, P. J. 1664. *Muhammedismus in angelis Euphrataeis S. Johanni Apocal. IX. v. XIII. & seq. praemonstratus*, Straßburg: Literis Georgi Andreae Dolhopfii.

SPENER, P. J. 2005. *Pia desideria*. Gießen: Brunnen-Verlag.

SZYRWIŃSKA-HÖRIG, A. 2024. Philipp Jacob Speners Einfluss auf Kants Auffassung von der Erbsünde. In: HANKE, T.; SANS, G. (eds.). *Kants Theorie des Bösen im Kontext*. Hamburg: Felix Meiner Verlag.

THEIS, R. 1994. *Gott*: Untersuchung zur Entwicklung des theologischen Diskurses in Kants Schriften zur theoretischen Philosophie bin hin zum Erscheinen der Kritik der reinen Vernunft. Stuttgart-Bad Canstatt: Frommann-holzboog.

"The use of the word 'nature' [...] is more befitting the limitations of human reason": Kant and the semantics of nature

"O uso da palavra 'natureza' [...] é mais adequado aos limites da razão humana": Kant e a semântica da natureza

Tania Eden
Ruhr-Universität Bochum
tania.eden@ruhr-uni-bochum.de

Abstract: The paper analyses selected passages from Kant's critical and pre-critical writings in order to articulate a unified framework that reconciles causal mechanism, divine purposiveness and teleology within the Kantian approach to practical philosophy.

Keywords: freedom; History; human reason; Kant; nature; semantics.

Resumo: O artigo examina trechos selecionados dos escritos críticos e pré-críticos de Kant a fim de promover uma análise unificada capaz de reconciliar mecanismo causal, intencionalidade divina e teleologia no âmbito da abordagem kantiana da filosofia prática.

Palavras-chave: liberdade; História; razão humana; Kant; natureza; semântica.

Recebido em 10 de setembro de 2025. Aceito em 06 de dezembro de 2025.

doispontos, Curitiba, São Carlos, vol. 22, n. 3, dez. de 2025, p. 33-41 / ISSN: 2179-7412
DOI: 10.5380/dp.v22i3.101184

What is the driving force of human history? Is it the exercise of human freedom – our capacity to set ends according to reason and choose actions as means to achieve them? Or is history driven more by a kind of Newtonian nature – by causal mechanisms that operate independently of human reason or will? A variation of the second answer could be based on Kant's pre-critical perspective. In his *Universal Natural History and Theory of the Heavens* (1755), Kant argues that God's goodness and omnipotence are demonstrated not through miracles or supernatural interventions, but rather through the lawful and mechanical order of the natural world[1]. Applying this framework to human history suggests that the unfolding of historical events could similarly be understood as part of a larger network of causes and effects, progressing in accordance with a divinely ordained providence. This introduces a problem that goes beyond the classical problem of theodicy, namely the question of how mechanistic explanations of nature can coexist with divine purposiveness and teleology:

> In the arrangement of the solar system [...] we have recognized the wisdom of God which has so beneficially ordered everything for the good of the rational beings that inhabit them. However, how can one now reconcile a mechanical doctrine with the teaching of intentions in such a way that what the highest wisdom itself designed has been delegated for implementation to coarse matter and the regiment of providence to nature left to its own devices? [...] Must not the mechanics of all natural motions have an essential tendency to many such consequences that accords with the project of the highest reason in the whole extent of connections? (1:363)

A Newtonian view of nature seems to leave no room for anything resembling a "teleological doctrine of nature" (8:18). A universe operating according to mechanical laws, functioning as a closed and autonomous system, seems to require no divine creator and could, presumably, explain its existence from within itself (LALLA, 2003, 436). The young Kant recognizes this challenge:

> The defender of religion is concerned [...] that if natural causes can be discovered for all the order in the universe that can be brought about by the most general and most essential properties of matter, then it is not necessary to invoke a highest governing power (1:223).

Before addressing moral questions concerning divine providence, a theoretical problem inherent to any cosmogony must be resolved. How can an explanation of the universe, based solely on Newton's mechanical laws, be reconciled with God's purposeful plans for creation? Furthermore, how might a teleological account of nature be integrated with the natural laws governing our phenomenal world? A critical solution to this problem will be developed in Kant's mature work. If an appeal to teleology is to be legitimized and made the basis for an account of human history, it could only serve as a heuristic tool or regulative principle, helping to make nature more intelligible to us by satisfying the inevitable striving of human reason for systematic unity and overarching theories. However, a look at Kant's early writings proves useful for understanding his attribution of purposiveness to nature and his occasional preference for using the term 'providence' instead of 'nature' in later works.

Let us consider two examples. In the nineth proposition of the pivotal essay *Idea for a Universal History with a Cosmopolitan Aim* (1784) Kant refers to his view of history as a "justification of nature", immediately qualifying his choice of terminology with a parenthetical remark:

> Such a justification of nature – or better, of providence – is no unimportant motive for choosing a particular viewpoint for considering the world. For what does it help to praise the splendor and wisdom of creation in the nonrational realm of nature, and to recommend it to our consid-

[1] Kant's writings will be cited below according to the Berlin Academy Edition of Kant's Collected Works, with volume and page numbers provided. For the English translation see The Cambridge Edition of the Works of Immanuel Kant.

> eration, if that part of the great showplace of the highest wisdom that contains the end of all this – the history of humankind – is to remain a ceaseless objection against it, […] (8:30).

And in *On the Common Saying: This May Be True in Theory, but It Does Not Apply in Practice* (1793), Kant points out that progress in history could not be achieved if nature did not already steer us in a particular direction – or, as Kant puts it, – "force us onto a track we would not readily take of our own accord" (8:310). To convey this idea of enforcement, Kant once again emphasizes that 'providence' might be the more appropriate term:

> If we […] ask by what means this unending progress toward the better can be maintained and even accelerated, it is soon seen that this immeasurably distant success will depend not so much upon what we do […] and by what methods we should proceed in order to bring it about, but instead upon what human nature will do in and with us to force us onto a track we would not readily take of our own accord. For only from nature, or rather from providence (since supreme wisdom is required for the complete fulfillment of this end), can we expect an outcome that is directed to the whole and from it to the parts, […] (8:310).

Let us explore the context of this remark. In On the Common Saying Kant responds to contemporary debates on the possibility of cultural and moral progress. He quotes Moses Mendelssohn, who argued that while individuals may excel and achieve great things, humanity as a whole pays for each advancement with setbacks and regressions. According to Mendelssohn, there is no such thing as species-wide progress in human history. To counter this pessimistic view of humanity's future, Kant appeals to our duty to contribute to the development and perpetuation of human capacities through education and tradition.

> I rest my case on my innate duty, the duty of every member of the series of generations – to which I (as a human being in general) belong and am yet not so good in the moral character required of me as I ought to be and hence could be – so to influence posterity that it becomes always better (the possibility of this must, accordingly, also be assumed), and to do it in such a way that this duty may be legitimately handed down from one member [in the series of] generations to another (8:309).

The clause in parenthesis is essential. Just as each of us has a duty to strive for moral self-perfection, we also have a duty to assist future generations in developing their rational and moral faculties through education and critical self-reflection. Since it is both pointless and normatively ineffective to impose a duty without a corresponding ability in those obligated, we must assume that moral cultivation and improvement over time are indeed possible. Ought implies can. This principle is well known from Kant's critical works. When applied to an individual's moral character, Kant emphasizes that no matter how evil a person's past actions may have been,

> his duty to better himself was not only a duty in the past; it remains his duty now. Therefore, he must be capable of it, and if he fails to act accordingly, he is, at the moment of action, just as accountable and stands just as condemned as if […] he had only just fallen from innocence into evil (6:41).

A pessimistic view of past moral progress does not absolve us from the duty to improve ourselves or to do our best for the moral education of young people. Instead, the burden of proof shifts: to argue that historical progress is impossible, Mendelssohn would need to present compelling reasons to challenge a fundamental principle of deontic logic.

After having argued his thesis regarding historical progress, Kant asks "by which means" progress is maintained and accelerated (8:310). His claim that progress in historical development relies less on our actions and deliberate choices and more on what human nature does "in and with us" (ibid.) ties back to Kant's early cosmopolitan view, according to which only a federation of states with coercive powers to enforce its laws can enable humanity to fulfill its rational and

moral capacities.

In the introductory remarks to the *Idea for a Universal History with a Cosmopolitan aim*, Kant acknowledges that humanity's "doings and refrainings on the great stage of the world" (8:17) appear to defy rational meaning. As the "collective result of people's free actions" (WOOD, 2009, p. 112), history seems to lack any discernible order or direction. However, Kant contends that there is a way to render "this nonsensical course of things human" (8:18) theoretically intelligible: we must try to understand human history as being guided by an "aim of nature" <*Naturabsicht*> (8:17). Nature has not only endowed us with predispositions for the use of reason but also guides their development in a teleological process, using specific means to achieve its end.

In the fourth proposition of the *Idea for a Universal History*, Kant maintains that progress in human history is partially driven by social antagonism, which he calls the "unsocial sociability" <*ungesellige Geselligkeit*> of men. While we have a "propensity to enter into society," we also possess a "thoroughgoing resistance" to socializing with others, making us prone to isolating ourselves and tearing society apart (8:20). Although Kant does not use the exact expression, this combination of opposing character traits has been aptly termed "the cunning of nature", alluding to Hegel's doctrine of "the cunning of reason" (YOVEL, 1980, p. 165). For if our social drives were not counterbalanced by self-serving and anti-social incentives, we would lack a crucial motivation to grow beyond ourselves and to enlighten and enrich the readily satisfied demands of daily life.

> Thanks be to nature, therefore, for the incompatibility, for the spiteful competitive vanity, for the insatiable desire to possess or even to dominate! For without them all the excellent natural predispositions in humanity would eternally slumber undeveloped (8:21).

Kant's most comprehensive account of social unsociability appears in his later work, *Religion within the Boundaries of Mere Reason* (1793). There are, Kant says, three original predispositions toward the good in human nature: animality, humanity, and personality. The predisposition to animality pertains to our biological being. Like unreasonable animals, humans possess a "physical or merely mechanical self-love" (6:26), which encompasses instinctive elements such as self-preservation, the sexual drive, and the desire for community. While the predisposition to animality serves the survival and flourishing of the human species, the self-love associated with humanity differs from this biologically rooted form. Here, reason becomes "subservient to other incentives" (6:28) and provides our sensual inclinations with the rational means to pursue happiness by achieving personal goals and self-chosen ends (cf. 5:73). Kant describes this rational capacity for self-love as one that is inevitably "comparative [...], although physical in nature" (6:27). Since the predisposition to humanity involves the rational faculty of comparison, we assess our well-being not only by evaluating and ranking potential objects and ends of deliberate choice, but also by comparing ourselves with others. "Out of this self-love originates the inclination to gain worth in the opinion of others" (6:27). We judge the lives we lead and the decisions we make based on the value they hold in the opinion of others, which provides a constant source of self-discontent. "[The human being] is poor (or considers himself so) only to the extent that he is anxious that other human beings will consider him poor and will despise him for it" (6:93). While our social nature inclines us to think of all people as equal, our unsocial drives spark suspicion that others might be better off than we are, leading to jealousy and rivalry.

> These vices, however, do not really issue from nature as their root but are rather inclinations, in the face of the anxious endeavor of others to attain a hateful superiority over us, to procure it for ourselves over them for the sake of security, as preventive measure; for nature itself wanted to use

> the idea of such a competitiveness [...] only an incentive to culture (6:27).

Reason is also central to the third predisposition, that of personality. Although humans are rational creatures, rationality alone does not inherently equate to morality (6:26, note). While the predisposition to humanity aligns with hypothetical imperatives, employing reason to pursue arbitrary, non-moral ends, the predisposition to personality "is rooted in reason that is practical in itself, i.e., in reason that legislates unconditionally" (6:28). Therefore, a third predisposition in human nature must be introduced – one that emphasizes our capacity to respect the moral law and to act in accordance with it, "by disregarding all desires and sensible incitements" (4:457).

Kant defends a teleological view of nature, stating that "everything in the world is good for something" (5:379). The first proposition of the *Idea for a Universal History* holds that nature intends all natural predispositions of an organic being to be developed completely and purposively. "An organ that is not to be used, an arrangement that does not attain to its end, is a contradiction in the teleological doctrine of nature." (8:18). Nature does nothing in vain; it adapts organisms to their environment so that their predispositions will, sooner or later, reach their full development.

Assigning an intrinsic purposiveness to nature is an indispensable part of our attempts to systematize knowledge of the empirical world. In the *Critique of Pure Reason* Kant argued that reason can arrive at the "greatest unity of experience" only when it is allowed, for heuristic purposes, to "also connect things according to teleological laws" (A 687/B 715). Viewing nature as if it were ordered by a supreme intelligence increases the intelligibility of nature, helping us to discover universal generalizations in experience and establish interconnections between them. However, applying teleological laws to nature does not introduce a supernatural ground of causality (5:383). The idea of natural purposiveness serves only as a regulative principle in our inquiries into nature, providing coherence and systematicity to our empirical cognitions, thereby extending the explanatory scope of Newton's mechanistic conception of nature. Kant is careful to insist that "the teleological way of judging" (5:382) complements but does not supplant physics or natural science. Searching for "a hidden plan of nature" (8:27) does not entitle us to explain empirical facts by reference to some superhuman intention.

In the *Critique of the Power of Judgment*, Kant explicitly distinguishes natural purposiveness from intentional purposiveness, emphasizing that he has chosen the expression "end of nature" <*Zweck der Natur*> precisely because he does not want to blur the boundaries between theology and science. As a matter pertaining to "the reflecting, not the determining power of judgment" (5:383), Kant's teleological doctrine of nature involves no metaphysical commitments:

> In teleology, insofar as it is connected to physics, we speak quite rightly of the wisdom, the economy, the forethought, and the beneficence of nature, without thereby making it into an intelligent being (since that would be absurd); but also without daring to set over it, as its architect, another, intelligent being, because this would be presumptuous <*vermessen*>; rather, such talk is only meant to designate a kind of causality in nature, in accordance with an analogy with our own causality in the technical use of reason, in order to keep before us the rule in accordance with which research into certain products of nature must be conducted (5:383).

One significant aim of Kant's critical philosophy, as developed in the Critique of Pure Reason, is "to deny knowledge in order to make room for faith" (B XXX). Faith is necessary because the laws of nature and those of morality are fundamentally independent of each other. Morality requires us to conceive of ourselves as members of an intelligible world that is not governed by

the causal determinism of Newtonian science. While the empirical 'freedom' of human beings is only relative, concerning the subjective pursuit of means for achieving sensible ends, transcendental freedom denotes an absolute freedom, which – without being compelled by sensible incentives or needing them at all – facilitates, under the command of moral law, the capability for "pure reason of itself" to be practical (5:121). Freedom of the will as a causal faculty means for Kant, "beginning a series of occurrences from itself, in such a way that in reason itself nothing begins, but as the unconditioned condition of every voluntary action, it allows of no condition prior to it in time" (B 581f.). Such a "causality of reason" does not occur in the causally determined world of appearances and can only be asserted as an "intelligible faculty" (B 579) inherent to the noumenal nature of humanity (6:239).

In the opening sentence of his 1784 essay, Kant alludes to his solution to the transcendental problem of free will, which he believes justifies assuming that our actions in the natural world are also free. He seeks to make the results of this freedom – our "doings and refrainings on the great stage of the world" – systematically intelligible by uncovering within them an unconscious, collective purposiveness, not guided by individual or collective choices, but by nature. Since teleological purposiveness, as a regulative idea, implies that something appears as if it were designed or produced according to the idea of some end or plan, two accounts of human history must be distinguished (cf. KLEINGELD, 2001). In the physical teleology of the philosophy of history, our observable actions in the empirical world must be understood as if they were destined to promote a peaceful and state-like federation of states as the natural end of humanity. In the moral teleology of the philosophy of religion, our inner freedom as members of the intelligible world must be understood as if it were destined to establish an "invisible church" (6:101) as the end of creation. We must consider the progress of humankind in its dual aspects: namely, "with respect to culture (as its natural end)" and "with respect to the moral end of its existence" (8:308/9). In both cases, we must assume that striving for an unattainable end deprives that end of any reasonable and motivating force. As Kant states in the *Critique of Practical Reason*, if the highest good that humans can achieve through their intelligible freedom – namely, to become not only worthy of happiness through moral virtue but also to experience happiness at some point in their lives – were impossible to realize, then the commands of moral law "would be fantastic, directed toward imaginary ends, and must therefore in themselves be false" (5:114).[2]

Against this background, it seems plausible to attribute Kant's choice of terminology to his intention to underscore the distinct tasks of the philosophy of history and the philosophy of religion. The personified use of 'nature' aligns with physical teleology and empirical history, while 'providence' pertains to faith and to what a good person may hope for, a concept reserved for the philosophy of religion. The title quotation from Kant's 1795 essay *Toward Perpetual Peace* corroborates this distinction between different lines of inquiry.

> The use of the word nature when, as here, we have to do only with theory (not with religion) is more befitting the limitations of human reason (which must confine itself within the limits of possible experience with respect to the relation of effects to their causes) and more modest than is the expression of a providence cognizable for us, with which one presumptuously puts on the wings of Icarus in order to approach more closely the secret of its inscrutable purpose (8:362).

In his detailed study of Kant's concept of providence, Ulrich Lehner describes a development that begins with Kant's preference in his early writings for a sapiential rather than an actualistic

[2] On Kant's stance on divine providence in relation to the highest good, cf. HAHMANN, 2016.

view, and that ultimately leads to an increasing marginalization of divine providence. From a critical perspective, nothing can be known about the relation of God to the world. Given the limitations of theoretical reason and our ignorance regarding things in themselves, Kant contends that reason does not provide us with a theoretical warrant for belief in divine providence. Instead, reason offers a practical warrant, according to which divine is to be restricted to the 'seeds' implanted in us, namely to our rational predispositions toward the good and the accompanying anti-social drives. Lehner concludes that "Kant's concept of providence radically closes God off from the world, substituting in its stead the actions of human praxis, thereby reducing providence to a providentia socialis seu immanens" (LEHNER, 2007, 484).

Although Lehner's assessment is broadly accurate, it remains incomplete in some important respects. The second proposition of *The Idea for a Universal History* states that the teleological process of the development of natural predispositions takes a different form for the human species. Unlike other natural creatures, whose predispositions are determined by instincts and confined to a single mode of life, human beings, as a rational species, have the capacity to lead a self-devised life and strive for self-improvement according to ends they adopt as their own (7:321). We must assume that human history is a learning process that can be promoted and encouraged through education, as well as through social and cultural institutions. "The human being can only become human through education. He is nothing except what education makes of him" (9:443). Human predispositions for the use of reason cannot fully develop within the lifetime of an individual but only gradually within the species as a whole, requiring an indeterminately long historical process (8:18). Relying on "a hidden plan of nature" and on divine providence as its source is not only a regulative idea aimed at unifying and systematizing our cognitions, but also of moral relevance, as it opens up "a consoling prospect of the future," (8:30) motivating us to improve ourselves and contribute to the betterment of posterity, even when our actual aspirations may fail. However, some caveats must be considered. Kant clearly envisages that empirical progress toward a legal order not only involves cultivation and civilization, but also moralization terminating in the transformation of the legal-political order into a "moral whole" (8:21).

To avoid depriving a virtuous person of the hope that this aim may one day be attained, it is surely not enough to advance the metaphysical hypothesis that God will, at the right moment, redirect the course of events – contrary to the very natural laws He Himself has established. On this matter, Kant is in agreement with Spinoza. Metaphysical hypotheses about the divine source of teleological purposiveness in nature miss the decisive point: what matters is the practical confidence in the effects of divine wisdom. However, Kant is well aware of the frailty of human nature. As humans are both sensible and rational beings, they are susceptible to sensual temptations, which generally makes it easier for them to follow inclination rather than duty. Due to his inherent predisposition - the 'physical' self-love, which he is not accountable for – man is not only attached to the incentives of sensuality but also finds the objects of his desires pressing upon him in an emotionally charged manner, thereby satisfying the empirical self "just as if it constituted our entire self" (5:74). Consequently, there arises the propensity "to make oneself as having subjective determinants of choice into the objective determining ground of the will in general" (ibid.). This propensity to elevate personal happiness to an "unconditional practical principle" of action results in turning self-love into "self-conceit" (ibid.), thus becoming a question of moral attitude.

Even though the "propensity to evil" (6:28) is inseparably connected with human nature, "ascent from evil back to the good" (6:45) is nevertheless possible. Religion has to do with what a good person may reasonably hope for. And in this context, Kant concedes that "some supernatural cooperation is also needed to his becoming good or better, whether this cooperation only consist in the diminution of obstacles or be also a positive assistance" (6:44). However, "the human being must nonetheless make himself antecedently worthy of receiving it" (ibid.).

Bibliographic References

FÖRSTER, E. 1998. Die Wandlungen in Kants Gotteslehre. Zeitschrift für philosophische Forschung, v. 52, n. 3, p. 341-362.

HAHMANN, A. 2016. Kants kritische Konzeption der Vorsehung im Kontext der Diskussion des höchsten Gutes. *Archiv für Begriffsgeschichte*, v. 58, p. 111-129.

KANT, I. -. *Cambridge edition of the works of Immanuel Kant*. Paul Guyer and Allen W. Wood (Eds.). Cambridge: Cambridge University Press.

KANT, I. 1900-. *Gesammelte Schriften*. Königlich Preussische Akademie der Wissenschaften (and successors) (Eds.). Berlin: de Gruyter (and predecessors).

KLEINGELD, P. 2001. Nature or providence?: on the theoretical and moral importance of Kant's philosophy of history. *American Catholic Philosophical Quarterly*, Minnesota, v. 75, n. 2, p. 201-219.

LALLA, S. 2003. Kants "Allgemeine Naturgeschichte und Theorie des Himmels" (1755). *Kant-Studien*, [*s.l.*] v. 94, p. 426-453.

LEHNER, U. 2007. Kants Vorsehungskonzept auf dem Hintergrund der deutschen Schulphilosophie und -theologie. In: *Brill's Studies in Intellectual History*, v. 149. Leiden: Brill.

WOOD, A. 2009. Kant's fourth proposition: the unsociable sociability of human nature. In: Oksenberg, R. A.; SCHMIDT, J. (Eds.): *Kant's idea for a universal history with a cosmopolitan aim:* a critical guide. Cambridge: Cambridge University Press.

YOVEL, Y. 1980. *Kant and the philosophy of history*. Princeton: Princeton University Press.

The online state of nature: Kantian perspectives on freedom of expression, platform power and information disorder

O estado de natureza online: perspectivas kantianas sobre liberdade de expressão, poder das plataformas e desordem informacional

Tailine Hijaz[1]
Universidade Federal do Paraná (UFPR)
tailinehijaz@hotmail.com

Joel T. Klein[2]
Universidade Federal do Paraná (UFPR)
jthklein@yahoo.com.br

Abstract: This article argues that the current system of digital platform governance amounts to an online state of nature, a condition characterized by provisional rights, the absence of public guarantees, and unilateral control over speech. Drawing on Kant's political philosophy, we examine how the concentration of power within digital platforms, exercised without subjection to public law, transforms the conditions under which civil freedom and meaningful public discourse can flourish. We relate this structure to the dynamics of information disorder, and argue that freedom of expression in the digital age must be rethought in light of public norms grounded in principles shared by all.

Keywords: state of nature; Kant; freedom of expression; digital platforms.

Resumo: O artigo argumenta que a configuração atual da governança das plataformas digitais constitui um estado de natureza online, marcado por direitos provisórios, ausência de garantias públicas e controle unilateral sobre o discurso. Com base na filosofia política de Kant, examinamos como o poder concentrado nas plataformas, exercido sem subordinação ao direito público, transforma as condições nas quais a liberdade civil e o debate público significativo podem ocorrer. Relacionamos essa estrutura às dinâmicas da desordem informacional e sustentamos que a liberdade de expressão na era digital precisa ser repensada com base em normas públicas fundadas em princípios compartilháveis por todos.

Palavras-chave: estado de natureza; Kant; liberdade de expressão; plataformas digitais.

[1] PhD Student at Federal University of Paraná (UFPR/Brazil) and Universität Vechta (Germany). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Project CAPES/DAAD/PROBRAL.

[2] Professor of UFPR/CNPq and Visiting Research Professor at Goethe University Frankfurt. This work is supported by ERC grant (LSR, Grant agreement number: 101170288).

Recebido em 30 de julho de 2025. Aceito em 11 de dezembro de 2025.

Introduction

In recent decades, digital transformation has profoundly altered the conditions under which information is produced, exchanged, and contested. As digital platforms have become central infrastructures for communication and interaction, they have accumulated significant political and economic power, often exercised without transparency or public accountability. These developments have also generated structural tensions between the principle of freedom of expression and the realities of online discourse, in which decisions about visibility, access, and communication are shaped by private actors operating outside public procedures.

In this article, we adopt a normative philosophical approach to this configuration. Drawing on Kant's concept of the state of nature, understood as a critical device rather than a historical hypothesis, we examine contexts in which there is no public law and the exercise of freedom depends on unilateral will and contingent force. We argue that Kant's distinction between provisional private and public rights provides a framework for evaluating the legitimacy of such arrangements. We further argue that major digital platforms within constitutional societies operate as pockets of a state of nature, functioning as normative enclaves where users, content, and algorithms interact under private rules and commercial incentives rather than publicly justified norms. In such spaces, the structural conditions for civil freedom and meaningful public discourse are not secured by law, but are instead left to the discretionary governance of private entities. From a Kantian perspective, this configuration cannot be regarded as morally legitimate.

The article is divided into three parts. First, we reconstruct Kant's account of the state of nature and the moral duty to establish public law, identifying the criteria that distinguish legitimate from illegitimate authority. Second, we develop a diagnostic analysis of platform governance, emphasizing how the regulation of speech is shaped by opaque, profit-oriented infrastructures. Third, we examine the normative implications of this configuration, and argue that freedom of expression in the digital age must be rethought in light of public norms grounded in principles shared by all.

1. The immorality of the state of nature: A Kantian framework

According to contractualist theories, civil society may have arisen naturally in historical terms, but it cannot be legitimized in a normative sense. In contrast to accounts that regard it as naturally determined, such as those of Aristotle, Hume, and Hegel, although from different perspectives, contractualist theories assume that the establishment of the state and the law requires specific justification. This entails providing criteria for distinguishing legitimate institutions from illegitimate ones. It is within this context that the concept of the state of nature emerges (KERSTING, 1994).

The concept of the "state of nature" offers a counterpoint to how human life would be organized if there were no state to create and guarantee law. Since the focus is on the state's legitimacy rather than its creation, the state of nature does not refer to a primitive state prior to the emergence of the state. Instead, it functions as a normative and conceptual construct for evaluating political institutions and the normative issues that arise in their absence. There are several concepts of the state of nature, and contractualist philosophers justify their understanding of law and justice based on how this concept is defined.

Within the framework of contractualist theories, Kant's model of the state of nature is particularly valuable for addressing the concept of a just state and for evaluating political and legal institutions. Kant seeks to construct a model that preserves certain elements found in the theories of Hobbes, Locke, and Rousseau while avoiding their respective shortcomings. From Hobbes, Kant retains the idea that, in the state of nature, individuals would live in constant physical and legal insecurity, since they could rely only on their own strength to enforce their claims. In such a condition, violence prevails, and individuals remain in perpetual tension and potential conflict with one another — perhaps not in an actual state of war, as Hobbes maintains, but rather in a condition of ever-present risk of conflict. Conversely, Kant rejects Hobbes's view that the state of nature entails a complete absence of sociability. From Locke, he adopts the view that rights can indeed be acquired in the state of nature, but he denies that such rights are absolute or that the civil state arises merely to preserve them. From Rousseau, he incorporates the idea that the social contract must reflect the will of everyone involved, understood as the general will in contrast to the will of all, while rejecting the notion that the content of the contract should be culturally or socially grounded in any particular conception of common happiness.

Some of the issues that Kant identifies in the models of the state of nature proposed by Hobbes, Locke, and Rousseau can be summarized as follows. Hobbes depicts the state of nature as purely individualistic. In this condition, even the strongest individual would remain vulnerable to violent death, for instance while sleeping. Yet a more plausible version of the state of nature would be one in which there exists a limited form of sociability. In such a case, individuals would gather in groups bound by loyalty and exercise their arbitrary will over others. Although still a condition of potential conflict, the state of nature is more accurately described as a struggle among more or less organized groups than as a struggle among isolated individuals. The model of Locke recognizes the social dimension of the state of nature but portrays it as excessively harmonious. He maintains that individuals generally acquire rights through peaceful means, by mixing their labor with unpossessed things. However, the idea of labor as the normative foundation of property is highly problematic, not only historically and culturally, but also philosophically, since it presupposes a direct and dogmatic relation between the person and the thing. Moreover, the philosophy of Locke offers no political or juridical remedy when this provision is not respected, giving his theory an intrinsic tendency to preserve the status quo. The model of Rousseau, in turn, presents two versions of the state of nature, both of which are equally implausible. The original state of nature is characterized by largely independent individuals who, unlike in the model of Hobbes, seek to remain isolated and generally avoid contact. By contrast, the degenerate state of nature described in the *Second Discourse* is defined solely by domination and deception, leaving no room for rightful claims.[3]

In light of these models, it is worth considering the merits of the Kantian model of the social contract. Unlike Hobbes's model, the state does not possess the legitimacy to eliminate rights that could have been acquired in the state of nature, since it does not create rights *ex nihilo*. In other words, the state is inherently limited and therefore precluded from becoming authoritarian. Unlike Locke's model, the state does not merely serve as an instrument of property owners for maintaining a given status quo. Unlike Rousseau's model, the legitimacy of the state does not rest on the customs of a specific community or on any particular conception of happiness.

[3] On the role of the concepts of the state of nature and the social contract in Rousseau's work see Consani and Klein (2022).

Grounded in the idea of the *omnilateral will*, the Kantian state can regulate and qualify rights, organizing them systematically in accordance with the principle of equal freedom rather than according to any conception of happiness. In other words, the state has no intrinsic tendency toward authoritarianism (as in Hobbes), injustice (as in Locke), or paternalism (as in Rousseau). Rather, it is a state with a liberal legal structure and a republican ethos that enables everyone to exercise their rights as they see fit, provided that they do not infringe upon the rights of others.[4]

Furthermore, unlike earlier theories, Kant does not justify leaving the state of nature through conditional reasoning grounded in self-interest. For Hobbes, if one wishes to avoid violent death, one must also wish to leave the state of nature. For Locke, if one intends to preserve the rights acquired in the state of nature, one must enter the civil state to ensure their protection. For Rousseau, if one seeks to recover the freedom and happiness lost in the "true state of nature," one must leave the corrupt state of nature and take part in the social contract. For Kant, however, the transition is not based on conditional reasoning ("If I want x, then I must do y") but on an unconditional duty of practical reason: "I ought to enter the civil state". The state of nature should be abandoned not because it is undesirable for some other reason, however compelling, but because it is intrinsically immoral and unjust. In Kant's terminology, the obligation to leave the state of nature constitutes a *categorical imperativ*e rather than a hypothetical one.

Following this broader conceptual delimitation within contractualist theories, our attention now turns to a more detailed examination of Kant's concept of the state of nature. While not exhaustive, the following discussion outlines its main features.

a) *On the morality of the civil state and the immorality of the state of nature.* According to Kant,

> A condition that is not rightful, that is, a condition in which there is no distributive justice, is called a state of nature (*status naturalis*). What is opposed to a state of nature is not (as Achenwall thinks) a condition that is social and that could be called an artificial condition (*status artificialis*), but rather the civil condition (*status civilis*), that of a society subject to distributive justice. For in the state of nature, too, there can be societies compatible with rights (e.g., conjugal, paternal, domestic societies in general, as well as many others); but no law, 'You ought to enter this condition,' holds a priori for these societies, whereas it can be said of a rightful condition that all human beings who could (even involuntarily) come into relations of rights with one another ought to enter this condition (MM 06:306; see also 06:242).[5]

This passage indicates that a social condition can exist within the state of nature, governed by what Kant calls private law. The postulate of public law follows from the very legitimacy of private law:

> [...] when you cannot avoid living side by side with all others, you ought to leave the state of nature and proceed with them into a rightful condition, that is, a condition of distributive justice. - The ground of this postulate can be explicated analytically from the concept of right in external relations, in contrast with violence (*violentia*)" (MM 06:307)[6].

[4] We are aware that our claims contrasting Kant's contractualism with that of Hobbes, Locke, and Rousseau are controversial. However, since these claims do not play a fundamental role in the argument of this paper, and because a proper justification of them would require an independent paper, we will leave this issue here as a contextual introduction and address it more carefully on another occasion.

[5] All translations follow *The Cambridge Edition of the Works of Immanuel Kant* (Kant 1992ff), with pagination according to the Akademie Ausgabe (Kant (1900-)). The works quoted in this paper are: MM – Metaphysics of Morals; TP - On the common saying: That may be correct in theory, but it is of no use in practice; WO - What does it mean to orient oneself in thinking?; WA - An answer to the question: What is enlightenment?; CJ – Critique of the power of judgment.

[6] It should be noted that Kant's legal philosophy is concerned with the regulation of external freedom, that is, with the external relations among agents. The aim, therefore, is not to evaluate or determine the agents' intentions at the moment of action. What matters primarily is the action itself, as it can be perceived by a third-party observer. Intentions become relevant only in the

Recognizing private law within the context of the state of nature is important for at least two reasons. First, denying its validity would imply that individuals who wish to enter into relations of rights and duties with one another would lack the legitimacy to do so. This would amount to stripping them of their status as persons and reducing them to the level of things. Second, acquired rights in the state of nature cannot be regarded as either entirely legitimate or entirely illegitimate. If they were completely legitimate, the state would have no moral function but only an instrumental one, as in Locke's theory. If they were completely illegitimate, the state would become the sole source of all law and morality, as in Hobbes's model. Kant positions his view precisely within this middle ground:

> The first and second of these conditions can be called the condition of private right, whereas the third and last can be called the condition of public right. The latter contains no further or other duties of human beings among themselves than can be conceived in the former state; the matter of private right is the same in both. The laws of the condition of public right, accordingly, have to do only with the rightful form of their association (constitution), in view of which these laws must necessarily be conceived as public (MM 06:306).

Claiming that the matter of private law remains the same within public law means that if someone can acquire something as their own within the framework of private law then that right cannot be eradicated within the civil condition. However, this right is not absolute, as in Locke's theory, since the state has a moral obligation to guarantee "the rightful form of their association (constitution), in view of which these laws must necessarily be conceived as public". Using marriage as an example, this means that the state may legitimately establish criteria for entering into marriage, such as being of legal age, having mental capacity, and not being already married. Ultimately, the state regulates private rights according to the idea of the omnilateral will. In this sense, the state does not determine the content of the law, which remains the same as in private law; rather, it regulates its form. Thus, the state cannot dictate whom a person should marry or whether they should marry at all, but it can establish rules that those who wish to marry must follow. This ensures that marriage functions as a legitimate legal institution compatible with other state institutions according to the idea of a system of equal freedom. Civil regulation is necessary because private rights are essentially indeterminate in their limits and can easily come into conflict with one another, leading individuals to protect their claims through violence or unilateral coercion. Within the state of nature, only violence or sheer force can enforce ambiguous and indeterminate rights. That is why

> [g]iven the intention to be and to remain in this state of externally lawless freedom, men do one another no wrong at all when they feud among themselves; for what holds for one holds also in turn for the other, as if by mutual consent (uti partes de iure suo disponunt, ita ius est.) But in general they do wrong in the highest degree by willing to be and to remain in a condition that is not rightful, that is, in which no one is assured of what is his against violence (MM 06:307f.).

It is precisely because of the uncertainty surrounding the limits of each person's rights and the means by which they can be enforced that Kant holds that private law in the state of nature is merely provisional. "So only in a civil condition can something external be mine or yours" (MM 06:256). This means that

> [...] a unilateral will cannot serve as a coercive law for everyone with regard to possession that is external and therefore contingent, since that would infringe upon freedom in accordance with universal laws. So it is only a will putting everyone under obligation, hence only a collective general (common) and powerful will, that can provide everyone this assurance. - But the condition of being

evaluation of illegitimate actions, where they may serve as aggravating or mitigating factors. In cases where the action is lawful, however, the evaluation of intent plays no role.

> under a general external (i.e., public) lawgiving accompanied with power is the civil condition (MM 06:256).

b) *Different spheres of the state of nature*. The concept of the state of nature allows us to adopt a perspective of illegitimacy that does not depend on historical or genealogical explanation. According to Kant, we may overcome the state of nature in one sphere while remaining within it in another. Thus, a people may have overcome the state of nature and established a civil condition within a certain domain, yet still find themselves in a state of nature in relation to other peoples existing alongside them. In this sense, "a state, as a moral person, is considered as living in relation to another state in the condition of natural freedom and therefore in a condition of constant war" (MM 06:343)[7]. One could therefore argue that, from one perspective, a civil condition exists, while from another, it has not yet been achieved. With the advent of the internet and social media, one could further argue that a new sphere has emerged in which human relations are no longer organized according to the standpoint of civil freedom but rather according to that of natural, or even savage, freedom.

c) *Right as a relation among persons*. According to Kant, private law in the state of nature cannot arise from a direct relationship between persons and things, as Locke proposed. Rather, legal relations can be established only among persons, and these relations sometimes refer to things. Therefore,

> it is clear that someone who was all alone on the earth could really neither have nor acquire any external thing as his own, since there is no relation whatever of obligation between him, as a person, and any other external object, as a thing. Hence, speaking strictly and literally, there is also no (direct) right to a thing. What is called a right to a thing is only that right someone has against a person who is in possession of it in common with all others (in the civil condition) (MM 06:261).

Claiming something as one's own implies a relation to another person, which, to be legitimate, must satisfy the criteria of proportionality and universal reciprocity. Even when acknowledging the potential legitimacy of rights that arise within the state of nature, such provisional rights must still conform to the categorical principle of right, which states: "Act externally in such a way that the free use of your choice can coexist with the freedom of everyone in accordance with a universal law" (MM 06:231).

d) *The shortcomings of an ethical approach*. The moral problem inherent in the state of nature cannot be resolved simply by educating individuals or cultivating their moral character, because it

> is not experience from which we learn of the maxim of violence in human beings and of their malevolent tendency to attack one another before external legislation endowed with power appears, thus it is not some deed that makes coercion through public law necessary. On the contrary, however well disposed and law-abiding human beings might be, it still lies a priori in the rational idea of such a condition (one that is not rightful) that before a public lawful condition is established individual human beings, peoples and states can never be secure against violence from one another, since each has its own right to do what seems right and good to it and not to be dependent upon another's opinion about this. So, unless it wants to renounce any concepts of right, the first thing it has to resolve upon is the principle that it must leave the state of nature, in which each follows its own judgment, unite itself with all others (with which it cannot avoid interacting), subject itself to a public lawful external coercion, and so enter into a condition in which what is to be recognized as belonging to it is determined by law and is allotted to it by adequate power (not its own but an external power); that is, it ought above all else to enter a civil condition (MM 06:312).

[7] See also: "In this problem the only difference between the state of nature of individual human beings and of families (in relation to one another) and that of nations is that in the right of nations we have to take into consideration not only the relation of one state toward another as a whole, but also the relation of individual persons of one state toward the individuals of another, as well as toward another state as a whole" (MM 06: 343f.).

In a civil condition, individuals do not need to rely solely on their own judgment about what is fair or unfair, because external coercive laws define these boundaries. Even if ethical education could protect us from all the negative consequences of the state of nature, such a condition would still be unjust, since it does not clearly specify the rights and duties of each individual from the systematic standpoint required by civil freedom. People could disagree in good faith about the definition and application of rights and duties, which means that there would be too much uncertainty to ensure the protection of anyone's rights.

However, as the modern state has evolved and political as well as legal relations have become more complex, the concept of the state of nature must be updated to address these new challenges. We propose the notion of pockets of the state of nature. By this, we mean contexts in which human relations are subject to the demands of rights and duties, since individuals can harm one another's external freedom without their actions being regulated by laws established by a competent authority. In such cases, agents may violate the freedom of others, and each must enforce their own rights. Although claims to rights and duties arise naturally from practical reason, these rights remain provisional, indeterminate, and uncertain, leaving room for unilateral interpretation by each agent.

Unlike the general concept of the state of nature, which functions as a conceptual counterpart to the civil condition, the idea of pockets of the state of nature presupposes the existence of a functioning civil state. It depends on the civil state because it requires an effective legal framework for such contexts to emerge. These pockets may arise naturally as new forms of external freedom relations develop, or they may result from deliberate actions by groups that seek to benefit from a civil state guaranteeing a legal superstructure while simultaneously withdrawing from or restricting its application in certain domains. This leaves individuals to determine their rights unilaterally within those spheres.

From a Kantian point of view, even if the state has not created laws to regulate a particular domain, all human relations capable of affecting the external freedom of others must remain open to possible regulation. For such regulation to be legitimate, however, it must proceed according to the idea of an omnilateral will rather than a unilateral or even multilateral one, since various powerful agents may otherwise unite to impose their interests on weaker parties.

Given this context, the next section will examine whether digital platforms reproduce, in new forms, a condition analogous to the state of nature. This condition is marked by the absence of public guarantees, by power asymmetries, and by the indeterminacy of rights. Building on this hypothesis, we will explore how digital platforms contribute to the current crisis of informational disorder.

2. Platforms as private governors: diagnosing the crisis of information disorder

The internet has radically transformed the conditions under which information is produced, circulated, and consumed. It has given rise to new forms of collective intelligence (LÉVY, 2015)[8], operationalized "global networks of instrumentalities" (CASTELLS, 2018, p. 77), and inaugurated

[8] According to Lévy, collective intelligence "is a distributed intelligence everywhere, continuously enhanced, coordinated in real time, which results in the effective mobilization of skills". He adds that "the basis and goal of collective intelligence are the mutual recognition and enrichment of people, not the cult of fetishized or hypostatized communities" (LÉVY, 2015, p. 29). Translated by us from: "é uma inteligência distribuída por toda parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva das competências" and "a base e o objetivo da inteligência coletiva são o reconhecimento e o enriquecimento mútuos das pessoas, e não o culto de comunidades fetichizadas ou hipostasiadas".

a new "Galaxy" distinct from traditional mass media (CASTELLS, 2003)[9]. This shift not only expanded access to the public sphere but also reshaped the architecture of expression in notable ways. However, it also led to the concentration of power in the hands of a few platforms — including Meta (Facebook, Instagram, WhatsApp), Google (YouTube), X (formerly Twitter), and TikTok — whose roles now exceed those of neutral intermediaries and who wield significant influence over the digital public sphere[10]. For instance, Facebook alone reported nearly 3.43 billion daily active users worldwide as of March 2025, more than twice the population of China[11]. Moreover, their annual revenues surpass the GDPs of numerous nations, and their infrastructures mediate everyday forms of social interaction, information access, and political deliberation on a global scale.[12]

As Klonick observes, these companies no longer operate as neutral conduits for third-party content; instead, they have become "New Governors", private entities that self-regulate based on a mix of economic interests and perceived alignment with democratic expectations (2018, p. 1603). Rather than merely hosting content, they act as norm entrepreneurs, interacting with users, states, and other stakeholders, apart from making discretionary decisions about what can be said, shared, or silenced. Indeed, Klonick describes how they establish internal rules, maintain enforcement mechanisms, and act through centralized authority (2018, p. 1663). A well-known example is the suspension of Donald Trump's account (@realDonaldTrump) after the January 6 Capitol attack. Although the correctness of this decision can be debated, it was implemented through private policies rather than judicial orders or constitutional adjudication[13].

While these actors now perform roles that were once reserved for public institutions, they remain private corporations driven by user engagement and profit maximization rather than public reason or constitutional duties. Unlike democratic governments, which are, in principle, subject to transparency, justification, and equal treatment, social media platforms prioritize the maximization of user attention, a practice that is often monetized through targeted advertising

[9] Castells presents the example of YouTube to clarify that, although it is a mass communication medium, it is distinct from traditional media. Due to the interactivity and horizontality of networks, "anyone can post a video on YouTube, with some restrictions. It is the user who selects the video they want to watch and comment on from a vast list of possibilities" (CASTELLS, 2018, p. 21). Translated by us from: "qualquer um pode postar um vídeo no Youtube, com algumas restrições. É o usuário que seleciona o vídeo que quer ver e comentar a partir de uma enorme lista de possibilidades"

[10] See Persily (2022, p. 200), who points out the complex nature of these platforms, arguing they are more than passive intermediaries like "common carriers", highlighting the role of algorithms in organizing content, and arguing that traditional legal frameworks developed for government speech suppression do not adequately apply to them: "[t]hey are not merely hosting speech, but organizing it. The most important feature of the platforms is the algorithms they employ to structure a unique 'feed' for every individual user".

[11] See Meta Reports First Quarter 2025 Results here: https://s21.q4cdn.com/399680738/files/doc_news/Meta-Reports-First-Quarter-2025-Results-2025.pdf.

[12] While revenue and GDP measure different things, the comparison remains valid to illustrate the scale of these corporations' economic power.

[13] As explains Franks (2022, p. 75), "Twitter had first temporarily locked the @realDonaldTrump account on January 6 after Trump posted a video and a statement repeating false claims about the election and expressing his 'love' for the rioters, requiring Trump to delete the tweets before being able to post again. At the time of the lockout, the Twitter Safety team noted that if Trump violated Twitter's policies again his account would be banned. In a blog post on January 8, the company explained that it had determined that two of the Trump tweets that followed the riots, one referencing 'American Patriots' and another stating that Trump would not be attending President- Elect Biden's inauguration, were 'likely to inspire others to replicate the violent acts that took place on January 6, 2021, and that there are multiple indicators that they are being received and understood as encouragement to do so'".

and behavioral data extraction (ZUBOFF, 2019, pp. 31, 94, 97, 115)[14]. Moreover, in this system, virality frequently supersedes veracity, and content designed to provoke outrage or entertain tends to outperform content intended to inform or foster deliberation. Algorithmic curation also prioritizes emotional intensity and interaction volume (ZUBOFF, 2019, pp. 137–138, 580), often diminishing the visibility of reasoned or meaningful contributions to public debate[15] [16].

For these reasons, the economic architecture of networks fosters the spread of problematic content, such as misinformation, disinformation, and bullshitting. Despite the frequent conflation of these concepts, they refer to distinct practices. In short, misinformation is defined as the unintentional propagation of false content, whereas disinformation signifies the deliberate dissemination of false or manipulated content intended to attain political or economic ends (EUROPEAN COMMISSION, 2018; ORGANIZATION OF AMERICAN STATES, 2019)[17]. Bullshitting, as theorized by Frankfurt (2005) and elaborated by Cassam (2019, p. 80), refers to communicative behavior characterized not by deception but by disregard for truth. In reality, the bullshitter disregards factual accuracy and focuses primarily on persuasion or impression management. Therefore, in an environment as described, where visibility and virality are rewarded, all these forms of content find fertile ground.

Another challenge that must be addressed is the sheer volume of content. Indeed, the sheer scale of digital communication, when considered in conjunction with anonymity, bot activity, and algorithmic personalization, effectively undermines users' ability to discern reliable information and further complicates the already difficult task of moderating online content. As Wardle and Derakhshan (2017) note, distinctions between truth and falsehood are becoming increasingly indistinct, and emotionally charged or absurd claims can attain disproportionate visibility. These conditions erode shared epistemic standards and weaken the basis for meaningful disagreement and democratic legitimacy. Consequently, the collapse of factual consensus engenders cynicism, disorientation, and distrust in both traditional media and institutional authority.

On this topic, Vosoughi, Roy, and Aral (2018) demonstrated that false information spreads faster and more widely than true content, partly because it aligns with group identities and elicits more robust emotional responses[18]. Additionally, BuzzFeed News indicated that during the final months of the 2016 U.S. presidential election, the twenty most-engaged fake news stories

[14] This paper does not aim to explore the concept of "surveillance capitalism" but, for a detailed account, see Zuboff's *The Age of Surveillance Capitalism* (2019, part 1.3), which argues that this new economic order unilaterally claims private human experience as free raw material (what she terms "behavioral surplus") to produce prediction products sold in behavioral futures markets, creating unprecedented asymmetries of knowledge and power.

[15] On this point, Zuboff uses the example of professional journalism (whose purpose is to separate truth from falsehood) to show how Facebook's News Feed, for instance, treats all content alike, regardless of its truthfulness. According to Zuboff, in this context information disorder is treated as problematic unless it threatens business operations, and moderation is framed as a form of self-protection for the platform rather than public responsibility (2019, p. 138 [online]).

[16] For a comprehensive examination of the complexity of algorithms, see Bucher (2018), who conceptualizes them as sociomaterial assemblages shaped by code, people, and context, embedded with biases, and actively shaping social reality.

[17] For a more detailed discussion in the specialized literature regarding the definitions of fake news and disinformation, as well as a proposed legal definition of the latter, see subsection 2.2.3 in Hijaz (2023, pp. 125-138).

[18] "Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information. We found that false news was more novel than true news, which suggests that people were more likely to share novel information. Whereas false stories inspired fear, disgust, and surprise in replies, true stories inspired anticipation, sadness, joy, and trust" (VOSOUGHI, ROY, ARAL, 2018).

outperformed the twenty most popular news items from traditional media (SILVERMAN, 2016). These empirical studies confirm that platforms do more than merely host public discourse; they shape, filter, and often distort it through design choices and commercial incentives. Therefore, as discussed throughout this section, the information ecology inherently prioritizes content that is captivating over that which is accurate, and content that is polarizing over that which is reasoned.

In this context, the governance of the new digital public sphere is not guided by established public principles such as legality, transparency, due process, or other values deemed fundamental to a society that claims to be democratic. Instead, it is governed by opaque terms of service enforced through discretionary and often automated mechanisms. As Suzor observes, "[t]his is the opposite of the standards we expect of legitimate, legal decision-making in a democratic society" (2019, p. 8). Indeed, "when such companies make decisions about who uses their networks and how, they have almost unlimited discretion" and "[t]hey are accountable only to the market; there are no checks and balances on how they wield their power" (SUZOR, 2019, pp. 6–7). Furthermore, as noted in the introduction of this work, their compliance with public law tends to be reactive and selective, driven by reputational pressure or regulatory threats rather than institutional obligation. Consequently, fundamental decisions about visibility, access, and expression are made by actors who are not bound by constitutional duties and are subject to limited public scrutiny.

This phenomenon creates significant challenges in the enforcement of constitutional rights, including those related to freedom of expression, personal image, or honor against private entities. In instances where users contest the removal of content or the imposition of restrictions, their recourse is generally limited to internal appeals. Given the vast volume of content processed daily, moderation relies heavily on automated tools or rapid human review, restricting the scope for context or subtlety (ROBERTS, 2019, p. 179)[19]. Combined with a general lack of transparency, this frequently leads to mistakes. For instance, the iconic Vietnam War photo of the "Napalm Girl" was removed by Facebook due to its depiction of nudity, despite its historical significance, and was reinstated only after public backlash (LEVIN, WONG, HARDING, 2016). Conversely, extremely harmful content such as extremist videos has often remained available on platforms even after being reported. When removed, such content is frequently reinstated through re-uploads, often after having already been widely viewed and shared (LEVIN, 2017). Ultimately, as Keller (2019, p. 2) stresses, "users have few or no legal rights when platforms take down their posts".

When addressing the enforcement of fundamental rights in the digital sphere, one must consider the triangular dynamic between platforms, states, and users, a relationship that, although it varies by jurisdiction, also reveals common patterns. In the United States, for instance, the First Amendment restricts government action but does not obligate platforms to uphold speech rights internally[20]. Brazil's Civil Rights Framework for the Internet (Marco Civil da Internet) similarly

[19] We do not intend here to evaluate how good these decisions are or how they should ideally be made, but it is important to note that such decisions are made exclusively based on the platform's own standards of what constitutes free speech, acceptable discourse, and what does not. This complex situation creates an informational environment where users (and citizens) face insecurity, unpredictability, and vulnerability to arbitrary or inconsistent decisions, weakening the stability and fairness essential to a democratic public sphere (GORWA, 2019; FLEW; MARTIN; SUZOR, 2019).

[20] In this specific case, the outcome could even be seen as positive, considering that freedom of expression in the United States protects speech in cases where many would consider it reasonable to impose limitations, such as false speech, hate speech, and similar content. Nevertheless, it remains striking that a private platform ultimately defines what speech is acceptable.

grants platforms intermediary immunity, with exceptions limited to court-ordered takedowns[21] [22]. While a singular cause for this phenomenon remains elusive, the absence of clearly delineated and consistently applied criteria to guide enforcement mechanisms and internal decision-making processes within major platforms has given rise to competing claims among scholars. On the one hand, some argue that platforms overstep by censoring dissent[23]; on the other, critics contend that they fail to prevent the spread of harmful content[24].

Indeed, some initiatives have attempted to address this governance deficit. Meta's Oversight Board, launched in 2020, is often cited as a step toward accountability. Structured as a quasi-judicial body[25], the Board reviews a limited number of moderation cases and issues binding decisions. It is composed of independent experts and financed by a trust endowed with $150 million by Meta, with the stated purpose of operating independently from the company[26]. However, the scope of its jurisdiction is limited, its policy recommendations are non-binding[27], and its role remains embedded within the same corporate infrastructure it is meant to oversee (KLONICK, 2020, p. 2464). Even in high-profile cases, such as the decision to uphold Trump's suspension, the Board's rulings often align with the company's prior choices, raising doubts about independence and institutional legitimacy (REICH; SAHAMI; WEINSTEIN, 2021, p. 215)[28].

Accordingly, the issue appears to extend beyond the absence of regulatory oversight. In practice, regulation already exists, but it is implemented by the platforms themselves, resulting in a form of public sphere governance that lacks legitimacy. In this arrangement, the scope of fundamental rights is effectively delegated to corporate actors whose priorities are primarily shaped by market

[21] According to article 19, "In order to protect freedom of expression and prevent censorship, internet application providers may only be held civilly liable for damages resulting from third-party content if, following a specific court order, they fail to take the necessary measures, within the scope and technical limits of their service and within the time frame established, to make the allegedly infringing content unavailable, except as otherwise provided by law" (Brazil, 2014). Translated by us from: "Com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de internet somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário".

[22] In contrast, the EU's Digital Services Act (DSA), adopted in 2022, imposes obligations on major platforms to proactively moderate content, address systemic risks, and facilitate user appeals. Despite the challenges associated with its implementation, including a potential lack of coverage of emerging technologies and decentralized networks, the DSA marks a significant shift toward stronger platform responsibility and regulatory dialogue. For a comprehensive overview of this topic, see G'sell (2023).

[23] See, for example, Dickinson (2022), who argues that dominant platforms such as Facebook, Google, and X exert excessive control over public discourse, increasingly acting as private regulators of speech and silencing controversial or unpopular voices under the guise of content moderation.;

[24] See Truong et al. (2025), who argue that when social media platforms compete for user engagement, stricter moderation policies may lead to user migration to less regulated spaces, creating disincentives for effective regulation of misinformation and resulting in a collective under-enforcement scenario.

[25] When addressing this topic, Reich, Sahami, and Weinstein chose the following subtitle: "A Supreme Court for Facebook"? (2021, p. 213).

[26] "In 2019, Meta (then Facebook) established an irrevocable trust and transferred $130 million for the set-up and operations of the Oversight Board to the Trustees. On July 22, 2022, Meta announced additional funding of $150 million to be transferred to the Trustees as part of a commitment to provide ongoing financial support to the Oversight Board" (OVERSIGHT BOARD, online).

[27] To clarify, the decisions are binding for the specific cases reviewed by the Board, but its general policy recommendations are not.

[28] "In late spring 2021, it upheld Facebook's decision to suspend Trump but rejected an indefinite ban. It gave Facebook six months to revisit the case and provide clear, public standards for any continuing ban. A Solomonic decision, it pleased no one and returned power to Facebook. And odd choice if Oversight Board was intended to diminish the unchecked power of Facebook in deciding the boundaries of permissible speech on its platform" (REICH, SAHAMI, WEISTEIN, 2021, p. 215).

incentives. Meanwhile, national legal systems grapple with the challenge of constraining the transnational reach of these platforms, and international regulatory initiatives remain incipient and weakly binding. We argue that this dynamic creates a regulatory vacuum, wherein critical decisions about fundamental values, expression, truth, and participation are made in the absence of the usual constraints or safeguards that govern analogous public functions. From a philosophical standpoint, it is therefore important to question the moral implications of entrusting such regulatory decisions to a select group of a few transnational corporations. Thus, building on the Kantian argument set forth in the first section — particularly the concept of pockets of a state of nature —, the following will explore the normative implications of this arrangement and argue that a certain form of public regulation could offer a more legitimate and morally adequate response.

3. The political and legal approach to justifying and limiting online platforms

3.1. Why consider the digital environment a pocket of the state of nature?

As discussed in the first part of this paper, Kant's definition of the state of nature is characterized by the exercise of power without public authorization, by uncertain rights, and by a freedom dependent on force or on the convenience of the parties involved. We hold that this description provides a valuable framework for diagnosing the evolving regulatory landscape of major digital platforms, where decisions about what can be said, seen, or removed are made unilaterally through opaque internal procedures and with limited adherence to public legal standards. For instance, companies such as X and Rumble adopt different responses to judicial orders depending on the country, revealing a selective logic guided by political, reputational, or economic interests rather than by the uniform application of law[29].

In this context, the hypothesis advanced here is that these platforms operate as pockets of a state of nature — spaces that exist within a formal legal order yet in which certain actors systematically resist the transition to a civil condition by retaining the autonomy to define their own norms. This hybrid — and convenient — configuration combines the benefits of state stability, such as legal infrastructure, economic systems, and contractual protections, with the preservation of unilateral freedom that is not subject to obligations of public justification. In other words, this mode of operation preserves the advantages of civil society while maintaining the absence of constraints characteristic of the savage freedom described by Kant.

As discussed in the second part, the internet has enabled this kind of arrangement: certain groups benefit from public infrastructure, yet whenever regulation conflicts with their interests, they reject state oversight of interactions within digital environments and claim for themselves the exclusive right to determine what they consider legitimate. In turn, the criteria that govern content visibility or removal are shaped by internal rules that are often enforced through automated systems or private bodies without transparency or institutional guarantees of contestation. Consequently, users find themselves subject to unstable norms and decisions that evade any notion of legal transparency, thereby undermining both the predictability and the equality essential to the exercise of freedom of expression.

Thus, from the Kantian perspective developed here, even if certain rights may arise in the digital context, such as the right to access information or to express oneself in widely used virtual spaces,

[29] See: Campos Mello (2024); Global Freedom of Expression (2024a; 2024b).

these claims must be evaluated in light of the categorical principle of right, which states: "Act externally in such a way that the free use of your choice can coexist with the freedom of everyone in accordance with a universal law" (MM 06:231). Kant conceives the state of nature not as a historical stage but as a condition defined by the absence of legitimate public authority and the indeterminacy of rights. In the context of digital networks, this means that while the state should not exercise material control over access to platforms, it can — and must — establish formal criteria and ensure that digital interactions conform to the same principles that govern other areas of law, such as the protection of minors, civil liability, and respect for fundamental rights. Insofar as they do not constitute a realm beyond public law, platforms must also be subject to the moral principles of right, guided by the elimination of unilateral coercion and by the public determination of rights and duties.

Therefore, we argue that the concept of *pockets of a state of nature* is particularly relevant in this context, precisely because overcoming such a condition does not entail creating a parallel order, such as a supposed digital civil state, which would merely perpetuate the state of nature among competing jurisdictions. Rather, practical reason requires that there be a single legitimate civil condition. In other words, the only way to overcome these pockets is to extend the principles of public law to the digital environment, thereby ensuring that all subjects are placed under the same legitimate authority.

3.2. Why focus on platforms rather than on individuals?

This work focuses on platforms because they exercise structural power in shaping the discursive environment. When analyzing the normative challenges posed by digital communication, it is essential to distinguish between individuals who express opinions and the agents responsible for structuring the conditions under which those opinions are formulated, circulated, and received. Ultimately, what can be said, by whom, with what reach, and with what consequences is determined by how platforms define usage policies, organize informational flows, and control mechanisms of visibility. Moreover, this power is not exercised through isolated acts but through the continuous management of the architecture of the digital public sphere.

Therefore, just as the state must, according to Kant, secure the conditions for the coexistence of individual freedoms without determining their content, it must also ensure that platforms enable expression by structuring public space without shaping it according to their own interests. Yet, by combining opaque algorithmic ranking criteria, nontransparent internal policies, and unilateral sanctions, platforms assume a role that goes beyond mere technical mediation, directly influencing the conditions under which discourse occurs.

In this context, the decision to analyze platforms as the central object arises from the recognition that the exercise of individual freedom depends on a minimally structured discursive space governed by public and accessible norms. When that space is organized by private actors wielding asymmetric power, individuals become dependent on decisions they neither control nor have the institutional means to contest. This asymmetry undermines users' external freedom by subjecting them to coercive practices that do not originate from an omnilateral will. For this reason, we argue that, from a normative perspective, regulating platforms requires establishing a set of institutional conditions that ensure freedom of expression is exercised as a form of public freedom.

3.3. Limiting the unlawful freedom of digital platforms as a condition for freedom of thought

Kant was one of the strongest defenders of freedom of thought. According to him, attempts by authoritarian states to restrict freedom of expression inflict profound harm on one of the most fundamental human rights, since freedom of thought is "the sole palladium of the people's rights" (TP 08:304). Freedom of thought is inseparably linked to freedom of expression. For Kant,

> The freedom to think is opposed first of all to civil compulsion. Of course it is said that the freedom to speak or to write could be taken from us by a superior power, but the freedom to think cannot be. Yet how much and how correctly would we think if we did not think as it were in community with others to whom we communicate our thoughts, and who communicate theirs with us! Thus one can very well say that this external power which wrenches away people's freedom publicly to communicate their thoughts also takes from them the freedom to think (WO 08:144).

Because of the intrinsic connection between freedom of thought and freedom of expression, any attempt to restrict or abolish the right to freedom of expression is always illegitimate, even if it were approved by all citizens, since it could never be reconciled with the omnilateral will. Kant addresses this issue in the context of religious debates, one of the major political and social challenges of his time. He was personally affected by this matter when he faced political persecution and the threat of censorship:

> May a people itself make it a law that certain articles of faith and forms of external religion, once adopted, are to remain forever? And so: May a people hinder itself, in its posterity, from making further progress in religious insight or from at some time correcting old errors? It then becomes clear that an original contract of the people that made this a law would in itself be null and void because it conflicts with the vocation and end of humanity; hence a law given about this is not to be regarded as the real will of the monarch, to whom counter representations can accordingly be made (TP 08:305).

The same reasoning applies to enlightenment, for which freedom of thought is likewise essential. In this regard, and echoing the passage above, Kant writes:

> [...] to renounce enlightenment, whether for his own person or even more so for posterity, is to violate the sacred right of humanity and trample it underfoot. But what a people may never decide upon for itself, a monarch may still less decide upon for a people; for his legislative authority rests precisely on this, that he unites in his will the collective will of the people (WA 08:39).

Thus, within the framework of public law, the state cannot decide on behalf of individuals what they cannot decide for themselves. The omnilateral will must express the demands of reason, and reason cannot consent to a rule that undermines the very conditions of its possibility[30].

Since the concept of freedom of thought has been established as a fundamental basis of the individual's subjective rights, it is necessary to inquire into the criteria for its realization. In this domain as well, submission to law is required, because the freedom of thought of one individual must be compatible with the freedom of thought of all others. Kant formulated this law through the distinction between the private and public uses of reason. The private use of reason occurs when an individual acts within a public role, broadly defined, which may include the standpoint of an official, an institutional representative, a citizen, or a person fulfilling a particular duty. The

[30] In this sense, Kant agrees with Rousseau's view that human beings cannot voluntarily renounce their freedom and become slaves. For both Rousseau and Kant, such a contract would necessarily be invalid, regardless of whether the parties consent to it. A central element of Kant's philosophy becomes relevant here: the morality of the categorical imperative and the concept of the omnilateral will depend on normative rather than factual conditions. In other words, the issue is not simply what people want, think, or do, but what they are entitled to want, think, or do as equal free beings. Thus, the question of political legitimacy does not concern the will of all, which is illegitimate and may be entirely partial. Rather, it concerns what people could will, provided that certain moral criteria are met to ensure the possibility of a general will (in Rousseau's terminology) or an omnilateral will (in Kantian terminology).

public use of reason, by contrast, refers to "that use which someone makes of it as a scholar before the entire public of the world of readers" (WA 08:37). The public use of reason requires that one offer reasons and arguments that others could accept, in accordance with what Kant calls the "maxims of common human understanding", formulated as follows: "1. To think for oneself; 2. To think in the position of everyone else; 3. Always to think in accord with oneself. The first is the maxim of the unprejudiced way of thinking, the second of the broad-minded way, and the third of the consistent way" (CJ 05:294).

According to Kant, the private use of reason must be limited, while its public use must always remain free. Thus, the same individual must make appropriate decisions to comply with certain rules based on the private use of reason. For example, as a driver, one must take the necessary steps to stop at a red light. Yet, at another time and under different circumstances, that same individual may argue that there should be no traffic light at that location, or even that traffic lights should be abolished altogether. In other words, as a passive citizen, the freedom to use reason privately must be limited, whereas as an active citizen, one must be free to use reason to argue for changing the legislation in question.

Often overlooked in the literature is the existence of an inverse proportionality in Kant's distinction between the private and public uses of reason: the greater the freedom of private use, the less the freedom of public use, and vice versa (see Klein 2023a, 2023b, 2015). If the freedom of the private use of reason is expanded beyond a certain point, such use no longer constitutes a use in which an individual acts within a public role, broadly defined, because there would no longer be any public role to follow, and one would find oneself in a state of nature. The state of nature is precisely characterized by an almost unlimited freedom of the private use of reason, in which there would be no space, or only a very fragile one, for the free exercise of the public use of reason. After all, how could one make free public use of reason if other individuals constantly restricted or undermined such use through the private exercise of their own reason? Conversely, the greater the space for the freedom of the public use of reason, the more the freedom of its private use must be restricted. If, for instance, we wish to have spaces where the very existence of God can be freely discussed and questioned, to invoke a fundamental issue that has long served as a banner for "holy wars", then the freedom of the private use of reason by believers must be restricted so that it does not undermine or limit such a space.

Therefore, we can conclude that for freedom of thought to exist, it is necessary to distinguish between the private and public uses of reason. The private use of reason should be subject to political and social constraints, whereas the public use of reason must remain entirely unrestricted. The only limits on the public use of reason are those imposed by the inherent standards of rationality and by the specific context in which it is exercised. Consequently, although the freedom of the public use of reason cannot be censored or placed under authoritarian control, it must still respect the rules that allow it to coexist with the public exercise of freedom by all others. These rules include the maxims of common human understanding, the principles of logic, and humanity's accumulated knowledge. Kant never elaborated on the precise normative content of these rules, perhaps because it varies depending on the context. For example, arguing as a citizen for changing a traffic regulation is one thing, while engaging in a debate about the safety and efficacy of a vaccine is quite another. What counts as a legitimate public use of reason, therefore, may vary from one context to another.

In this article, we will not develop the general rules governing the public use of reason in the digital context, as that will be left for another occasion. Here, we argue that digital platforms cannot be regarded solely as agents exercising the freedom of public reason, but rather as loci where individuals and institutions — each of which may be considered moral persons with rights and duties — exercise their freedom in both the private and public uses of reason. In this sense, platforms themselves engage in a private use of reason, which must be determined by the principles of the omnilateral will. In other words, platforms must function as spaces that make possible the coexistence of the public and private uses of reason by other agents. This does not deny digital platforms the right to uphold a particular worldview; however, they may do so only while respecting the rules of public reason. Failure to observe this limitation results in what we have referred to above as a pocket of the state of nature.

The actions of digital platforms, which represent their private use of reason, must be limited in order to uphold the freedom of thought of all users. The public function of platforms is to facilitate the free public use of reason by their users. Rather than determining the content of discourse, platforms should guarantee the conditions under which users can freely exercise their reason without coercion or manipulation. If they fail to do so, whether by promoting or tolerating a form of lawless or "savage" freedom, they should be subject to sanctions and, if necessary, removed from circulation. Just as an individual may be punished for committing a crime against another, digital platforms can be held accountable for providing a space in which the use of reason becomes incompatible with the categorical imperative of right. By preserving lawless freedom, digital platforms create a pocket of the state of nature in which users cause one another harm to the highest degree.

3.4. Why is education not sufficient to overcome the digital state of nature?

One could argue that information disorder might be resolved through virtual ethical education or civic training aimed at fostering user responsibility and critical engagement. Although such initiatives are highly relevant, they remain insufficient from a normative standpoint. As discussed in the first part of this work, the injustice of the state of nature lies not only in its empirical outcomes but also in its structural incapacity to establish reciprocal duties under juridical conditions. Therefore, the issue cannot be reduced to conditional reasoning such as "if people are educated, they will act critically and autonomously, and the problem will be solved", because the digital state of nature remains morally unacceptable even in the presence of ethically responsible individual behavior. Indeed, the absence of legitimate public authority and of a shared legal framework renders every right uncertain and every duty unstable, which is incompatible with the principles of civil freedom.

Even if an ethically educated virtual community could avoid some of the negative effects of the digital state of nature, the condition would still be unjust, since it fails to clearly establish reciprocal rights and duties in accordance with the systematic perspective required for civil freedom. At most, networks could reach agreements based on the will of all, which, for Rousseau, could be profoundly unjust and, for Kant, would remain a partial and contingent will. However, they would never satisfy the criteria of an omnilateral will, which is consistent with the categorical imperative of right. Any such agreement would be contingent, as it would fail to respect the intrinsic principles that regulate external freedom, such as publicity, political representation, and

the balance of powers among the different branches of the state. In short, it would not meet the structural criteria of the concept of a just state or of a republic.

Therefore, a digital sphere based solely on voluntary coordination or on initiatives led by the platforms themselves would not meet the requirements of civil freedom, at least not from the theoretical perspective adopted here. What is required is not merely ethical behavior within an already existing architecture, but a transformation of that very architecture so that it no longer depends on private discretion and becomes subject to principles that can be publicly justified and applied equally. Once again, we emphasize that education may help to reduce harm, but it cannot substitute for legitimate legal conditions. Ultimately, only a transition to a civil condition, through juridical extension and structural reform, can overcome the pockets of unilateral power that characterize the online state of nature.

4. Final Remarks: Toward a Civil Form of Online Freedom

According to a Kantian conception, freedom must be exercised under conditions that ensure reciprocal recognition among subjects and remain within the bounds of shared public norms. When asserted unilaterally, on the basis of force or self-interest, freedom cannot be regarded as legitimate. Applied to the digital environment, this requirement calls into question the current organization of platforms, which are structured by commercial incentives and sustained by decisions that are not subject to any form of public deliberation or public right. As we have sought to demonstrate, these companies have assumed regulatory functions, organizing the flow of information according to their own rules and concentrating power in ways that evade the requirements of publicity, justification, and equality that define the civil condition.

Therefore, we argued that the concept of online freedom should be reconsidered in light of a model that preserves individual autonomy while incorporating it into an institutional framework designed to safeguard equal access, normative predictability, and public justification. Moreover, we maintained that civil freedom is not defined by the absence of interference, but by the existence of a legitimate order that distinguishes between compatible and incompatible uses of freedom. This means that speech must circulate within boundaries that can be accepted by all participants, even in digital environments. In Kantian terms, the exercise of freedom requires a transition from the digital state of nature to a public order guided by shared principles and the concept of equal freedom.

In this context, the emphasis falls far more on the architecture of platforms than on the content of individual messages. It is also important to stress that the problem does not arise from the plurality of opinions, but from the asymmetric conditions under which that plurality is expressed. Therefore, structuring a space that enables the public use of reason requires publicly defined criteria for organizing visibility, moderation, and access. To reiterate, the aim is not to impose substantive truths — since everything must remain open to discussion in accordance with the freedom of the public use of reason — but rather to establish a shared institutional foundation. To *civilize digital freedom*, in this sense, means creating an order that protects expression rather than shaping it, ensures equal conditions rather than dictating content, and provides legal security rather than promoting discursive uniformity.

However, this transformation faces significant obstacles. For instance, the transnational nature of platforms makes it difficult to enforce public norms consistently, particularly when national

legal systems operate according to different scopes and values. Although initiatives such as the European Union's Digital Services Act represent steps toward the institutional governance of the digital sphere, they still lack global effectiveness and encounter political resistance. Furthermore, a persistent imaginary of absolute freedom on the internet, which in Kantian terms would correspond merely to an ideal of imagination that in reality amounts to nothing more than savage freedom, is sustained by both corporate interests and certain segments of public opinion, making it more difficult to develop an institutional conception of responsibility. Although philosophy does not offer direct solutions to these challenges, nor does it claim to do so, we have argued that it can contribute to the debate by providing normative criteria for assessing the legitimacy or illegitimacy of platform conduct in light of the principles of public right.

In the virtual context of digital networks, this means that the state cannot intervene to determine, in a general or material sense, who may or may not have access to platforms, but it can establish formal criteria for their use. For example, it may set a minimum age for joining a social network, as in a proposed bill in Australia that seeks to restrict access for minors, or define what distinguishes an offensive opinion from a hate crime. It is the role of the state to enact general laws that regulate digital interactions in accordance with the broader legal framework governing other domains of law. In short, since digital networks are not exempt from public law, they must also be subject to the moral principles of public right, which are grounded in the moral obligation to determine rights and duties and to eliminate unilateral violence in accordance with the principle of equal freedom.

Bibliographic References

BRAZIL. 2014. *Law no. 12,965 of April 23, 2014*, [Online]. Available at: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Accessed: June 2, 2025.

BUCHER, T. 2018. *If... Then:* Algorithmic Power and Politics. New York: Oxford University Press.

CAMPOS MELLO, P. 2024. Musk cumpriu centenas de ordens de remoção de conteúdo do X fora do Brasil sem acusar censura. *Folha de São Paulo.* Available at: https://www1.folha.uol.com.br/poder/2024/04/musk-cumpriu-centenas-de-ordens-de-remocao-de-conteudo-do-x-fora-do-brasil-sem-acusar-censura.shtml. Accessed: December 17, 2024.

CASSAM, Q. 2019. *Vices of the Mind.* From the Intellectual to the Political. Oxford: Oxford University Press.

CASTELLS, M. 2018. *A sociedade em rede*. Trad. Roneide Venancio Majer, 19ª ed. Rio de Janeiro/São Paulo: Paz e Terra.

CASTELLS, M. 2003. *A galáxia da internet:* reflexões sobre a internet, os negócios e a sociedade. Trad. Maria Luiza X. de A. Borges. Rio de Janeiro: Zahar.

CONSANI, C. F.; KLEIN, J. T. 2022. *Leituras de Rousseau*. Florianópolis: Nefiponline. Available at: https://nefipo.paginas.ufsc.br/files/2012/11/LEITURAS-DE-ROUSSEAU.pdf. Accessed: July 20, 2025.

CONSANI, C. F. 2015. Democracia e os discursos de ódio religioso: O debate entre Dworkin e Waldron sobre os limites da tolerância. *ethic@ - An international Journal for Moral Philosophy*, Florianópolis, v. 14, n. 2, pp. 174–197, dec. Available at: https://periodicos.ufsc.br/index.php/ethic/article/view/1677-2954.2015v14n2p174. Accessed: July 20, 2025.

DICKINSON, G. M. 2022. Big Tech's Tightening Grip on Internet Speech, [Online]. *Indiana Law Review*, v. 55, n. 1, p. 101–128. Available at: https://doi.org/10.18060/26418. Accessed: June 2, 2025.

EUROPEAN COMMISSION. 2018. *A multi-dimensional approach to disinformation:* report of the independent High level Group on fake news and online disinformation, [Online]. European Commission. Available at: https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation. Accessed: June 2, 2025.

FLEW, T.; MARTIN, F.; SUZOR, N. 2019. Internet regulation as media policy: Rethinking the question of digital communication platform governance. *Journal of Digital Media & Policy*, v. 10, n. 1, pp. 33-50. Available at: https://doi.org/10.1386/jdmp.10.1.33_1. Accessed: June 2, 2025.

FRANKFURT, H. G. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.

FRANKS, M. A. 2022. The Free Speech Industry, pp. 65-86. In: BOLLINGER, L. C.; STONE, G. R. (Editors). *Social Media, Freedom of Speech, and the Future of Our Democracy*. Oxford: Oxford University Press.

GLOBAL FREEDOM OF EXPRESSION. 2024a. Columbia University. *The Case of the X Ban in Brazil.* Available at: https://globalfreedomofexpression.columbia.edu/cases/the-case-of-the-x-ban-in-brazil/. Accessed: March 8, 2025.

GLOBAL FREEDOM OF EXPRESSION. 2024b. Columbia University. *The Case of the Rumble Ban in Brazil.* Available at: https://globalfreedomofexpression.columbia.edu/cases/the-case-of-the-rumble-ban-in-brazil/. Accessed: June 6, 2025.

GORWA. R. 2019. The platform governance triangle: conceptualising the informal regulation of online content. *Internet Policy Review*, n. 8, v. 2. DOI: 10.14763/2019.2.1407. Available at: https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content. Accessed: July 10, 2025.

G'SELL, F. 2023. The Digital Services Act (DSA): A General Assessment, [Online]. In: VON UNGERN-STERNBERG, A. (ed.). *Content Regulation in the European Union* – The Digital Services Act. Trier Studies on Digital Law: Verein für Recht und Digitalisierung e.V., Institute for Digital Law (IRDT). Available at SSRN: https://ssrn.com/abstract=4403433 or http://dx.doi.org/10.2139/ssrn.4403433. Accessed: June 3, 2025.

HIJAZ, T. 2023. *Quanto Vale a Liberdade?* O problema da desinformação entre os diferentes fundamentos da liberdade de expressão. São Paulo: Dialética.

KANT, I. *Cambridge Edition of the Works of Immanuel Kant*. Paul Guyer and Allen W. Wood (ed.). Cambridge: Cambridge University Press, 1996–.

KANT, I. *Gesammelte Schriften*. Ed. Königlich Preussische Akademie der Wissenschaften (and successors). Berlin: de Gruyter (and predecessors), 1900–.

KELLER, D. 2019. Who do you sue? State and platform hybrid power over online speech, [Online]. *Aegis Series Paper No. 1902*. Hoover Institution. Available at: https://www.hoover.org/research/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech. Accessed: June 2, 2025.

KERSTING, W. 1994. *Die politische Philosophie des Gesellschaftsvertrags*. Darmstadt: Wissenschaftliche Buchgesellschaft.

KLEIN, J. T. 2018. Kant on religious intolerance. *Philosophica* (Lisbon), v. 51, pp. 25–38. Available at: https://repositorio.ulisboa.pt/bitstream/10451/40702/1/JoelKlein_Philosophica_51.pdf. Accessed: July 20, 2025.

KLEIN, J. T. 2015. Freedom of the Press: A Kantian Approach. *Estudos Kantianos*, v. 03, p. 83-92, 2015. Available at: https://revistas.marilia.unesp.br/index.php/ek/article/view/5122. Accessed: July 20, 2025.

KLEIN, J. T. 2023a. Enlightenment as the normative principle of social rationality. *Studia Kantiana*, v. 21, p. 99-117. Available at: https://revistas.ufpr.br/studiakantiana/article/view/91982. Accessed: July 20, 2025.

KLEIN, J. T. 2023b. Liberdade versus irracionalidade acadêmica: uma análise a partir de um ponto de vista Kantiano. *Ethic@* (UFSC), v. 22, p. 691-716. Available at: https://periodicos.ufsc.br/index.php/ethic/article/view/95262/54867. Accessed: July 20, 2025.

KLONICK, K. 2018. The New Governors: The People, Rules, and Processes Governing Online Speech, [Online]. *Harvard Law Review*, v. 131, p. 1598–1670. Available at: https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf. Accessed: March 16, 2025.

KLONICK, K. 2020. The Facebook Oversight Board: Creating an Independent Institution to

Adjudicate Online Free Expression, [Online]. *The Yale Law Journal*, v. 129, n. 8, p. 2418–2499. Available at: https://www.yalelawjournal.org/feature/the-facebook-oversight-board. Accessed: March 16, 2025.

LEVIN, S.; WONG, J. C.; HARDING, L. 2016. Facebook backs down from 'Napalm Girl' censorship and reinstates photo, [Online]. *The Guardian*. Available at: https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo. Accessed: June 3, 2025.

LEVIN, S. 2017. Tech firms fail to stop abusive content – leaving the public to do the dirty work, [Online]. *The Guardian*. Available at: https://www.theguardian.com/technology/2017/dec/05/youtube-offensive-videos-journalists-moderators. Accessed: June 3, 2025.

LÉVY, P. 2015. *A inteligência coletiva:* por uma antropologia do ciberespaço. Trad. Luiz Paulo Rouanet, 10ª ed. São Paulo: Edições Loyola.

META. 2025. *Meta Reports First Quarter Results*. Available at: https://s21.q4cdn.com/399680738/files/doc_news/Meta-Reports-First-Quarter-2025-Results-2025.pdf. Accessed: June 6, 2025.

ORGANIZATION OF AMERICAN STATES. 2019. *Guia para garantir a liberdade de expressão frente à desinformação deliberada em contextos eleitorais*, [Online]. OAS. Available at: https://www.oas.org/es/cidh/expresion/publicaciones/DesinformacionElectoral.pdf. Accessed: June 2, 2025.

OVERSIGHT BOARD. *How Is the Oversight Board Funded?* [Online]. Available at: https://www.oversightboard.com/faq/. Accessed: June 3, 2025.

PERSILY, N. 2022. Platform Power, Online Speech, and the Search for New Constitutional Categories, pp. 193-212. In: BOLLINGER, L. C.; STONE, G. R. (Editors). *Social Media, Freedom of Speech, and the Future of Our Democracy*. Oxford: Oxford University Press.

RAWLS, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

REICH, R.; SAHAMI, M.; WEINSTEIN, J. M. 2021. *System Error.* Where Big Tech Went Wrong and How We Can Reboot. New York: HarperCollins.

ROBERTS, S. T. 2019. *Behind the Screen*: Content Moderation in the Shadows of Social Media. New Haven: Yale University Press.

SILVERMAN, C. 2016. This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook, [Online]. *BuzzFeed News*, November 16. Available at: https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook#.etwaV6WDZq. Accessed: May 28, 2025.

SUZOR, N. 2019. *Lawless:* The Secret Rules That Govern Our Digital Lives. Cambridge: Cambridge University Press.

TRUONG, B. T.; KIM, S.; NOGARA, G.; VERDOLOTTI, E.; SAHNEH, E. S.; SAURWEIN, F.; JUST, N.; LUCERI, L.; GIORDANO, S.; MENCZER, F. 2025. Delayed takedown of illegal content on social media makes moderation ineffective. Technical Report. *arXiv preprint*, arXiv:2502.08841. Available at: https://doi.org/10.48550/arXiv.2502.08841. Accessed: June 2, 2025.

VOSOUGHI, S.; ROY, D.; ARAL, S. 2018. The spread of true and false news online, [Online]. *Science*, v. 359, n. 6380. Available at: https://doi.org/10.1126/science.aap9559. Accessed: February 12, 2025.

WARDLE, C.; DERAKHSHAN, H. 2017. *Information Disorder:* Toward an Interdisciplinary Framework for Research and Policy Making [Online]. Council of Europe Report. Available at: https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html. Accessed: February 14, 2025.

ZUBOFF, S. 2019. *The Age of Surveillance Capitalism:* The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs.

Modo de pensar e progresso moral: considerações sobre o valor do caráter empírico à luz da *Religião nos limites da simples razão*

Mode of thought and moral progress: considerations on the value of the empirical character in light of Religion within the boundaries of mere reason

Nicole Martinazzo[1]
Universidade Federal do Paraná (UFPR)
nicole.martinazzo@gmail.com

Resumo: O presente artigo parte de uma análise da *Primeira Parte* da *Religião nos limites da simples razão* para pensar de que maneira Kant entende a articulação entre a revolução no modo de pensar e a mudança lenta e gradual nos costumes quando se refere ao aprimoramento moral dos indivíduos. Procura-se defender que, nesse caso, "reforma" e "revolução" correspondem a duas faces da mesma moeda: trata-se de duas maneiras de avaliar o mesmo processo de mudança, sob diferentes pontos de vista. Da impossibilidade de reconhecer com certeza a revolução da intenção [*Gesinnung*] de um sujeito, retira-se uma valorização do papel do desenvolvimento do caráter empírico na filosofia moral kantiana.

Palavras-Chave: Immanuel Kant; moral; *Gesinnung*; modo de pensar; caráter empírico; educação moral.

Abstract: This paper begins with an analysis of the *First Par*t of *Religion within the Boundaries of Mere Reason* in order to consider how Kant conceives the relation between the revolution in the mode of thought and the slow, gradual change in customs with regard to the moral improvement of individuals. It argues that, in this context, "reform" and "revolution" represent two sides of the same coin: they are two ways of assessing the same process of change from different perspectives. From the impossibility of ascertaining with certainty the revolution of a subject's intention [*Gesinnung*], one draws an appreciation of the role of the development of empirical character within Kant's moral philosophy.

Keywords: Immanuel Kant; morals; *Gesinnung*; mode of thought; empirical character; moral education.

[1] O presente artigo é fruto do desenvolvimento de parte de minha tese de doutorado, defendida no Departamento de Filosofia do Instituto de Filosofia e Ciências Humanas da Unicamp (financiamento FAPESP, processo 2018/01544-8). O argumento sofreu alterações significativas desde então, alterações estas que só foram possíveis pois contou-se com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Recebido em 27 de agosto de 2025. Aceito em 11 de dezembro de 2025.

Uma leitura que contemplasse somente os textos em que Kant se dedica à fundamentação de sua filosofia moral poderia levar o leitor a concluir que esta se reduz a um mecanismo que visa julgar as ações de um determinado indivíduo apenas enquanto tomadas isoladamente[2]. Entretanto, mesmo nesses textos ele mobiliza um conceito que será melhor explicado anos depois, na Primeira Parte da *Religião*, a saber, o conceito de *Gesinnung* (aqui traduzido como "intenção")[3]. Esse conceito é referido também, por vezes, pela noção de que os seres humanos possuiriam um "modo de pensar" (*Denkungsart*)[4].

Se a ênfase em uma instância que subjaz todas as escolhas de um determinado sujeito soluciona alguns problemas (ao evitar, por exemplo, certo atomismo moral)[5], ela pode criar outros a depender da maneira como é compreendida. Um dos problemas mais evidentes é a possibilidade de mudança do agente moral, visto que a Gesinnung representaria o fundamento comum de todas as máximas de um determinado sujeito e que Kant parte do pressuposto, em sua doutrina do mal radical, que tal fundamento é, nos seres humanos em geral, corrompido. Essa corrupção, veremos a seguir, advém de algo tão simples quanto a mera inversão das prioridades do agente, que passa a privilegiar outros incentivos em detrimento da lei moral.

Resta-nos a questão: se todo mal decorre da corrupção do fundamento de todas as máximas, como é possível ao ser humano tornar-se bom? Kant dedica a esse tema sua Observação Geral à Primeira Parte da *Religião*, texto no qual propõe que a restauração da predisposição para o bem nos indivíduos (e, portanto, sua transformação moral) deve ser entendida nos termos de uma "conversão moral" ou mesmo de um "renascimento" (RGV, AA 06: 47). Assim, o aprimoramento moral requereria uma *revolução* do fundamento de todas as máximas.

---

[2] A identificação das citações das obras de Kant será feita de acordo com a edição da Academia (AA), conforme o modelo: RGV, AA 06: 03. Em cada citação, encontra-se a abreviatura da obra, número do tomo e número da página. As abreviaturas seguem o título alemão e são as seguintes: IaG: *Ideia de uma história universal de um ponto de vista cosmopolita*; WA: *Resposta à questão: o que é o Esclarecimento*?; GMS: *Fundamentação da metafísica dos costumes*; KpV: *Crítica da razão prática*; RGV: *A Religião nos limites da simples razão*; MS: *Metafísica dos Costumes*.

[3] O termo alemão *Gesinnung* representa um desafio para os tradutores de Kant. A pluralidade de opções revela as dificuldades de tradução que esse termo suscita, como bem mostram os tradutores brasileiros da Doutrina da Virtude (ed. Vozes) na nota 5 (2013, p. 153) ao justificar sua escolha. "Intenção" (escolha de Monique Hulshof na *Crítica da razão prática* e do grupo de tradutores da Doutrina da Virtude), "atitude" (utilizada por Guido de Almeida em sua tradução da *Fundamentação da Metafísica dos Costumes*) e "disposição" (escolha de Artur Mourão ou sua variação "disposição de ânimo", utilizada por Bruno Cunha na recente tradução da *Religião nos limites da simples razão*) são opções que recobrem apenas aspectos do conceito. Por conta de tais dificuldades — e por se tratar de um artigo com uma seção dedicada a compreender o conceito de *Gesinnung* — ora manteremos o termo em alemão ao longo do texto, ora utilizaremos a tradução "intenção". Sendo assim, alteraremos a tradução dos textos citados quando necessário para garantir a consistência de nossa escolha. Por via das dúvidas, indicaremos entre colchetes o original em alemão. Opta-se por "intenção" sobretudo para deixar claro, também na língua portuguesa, a diferenciação entre *Gesinnung* (intenção) e *Anlage* (comumente traduzido por "predisposição" ou "disposição").

[4] Há passagens ao longo da *Religião* nas quais Kant equivale *Gesinnung* e *Denkungsart* (por exemplo, RGV, AA 06: 46n; 47) Talvez estas formulações não sejam suficientes para fixar uma equivalência geral entre eles, mas elas ao menos mostram que se trata de conceitos que em determinados contextos podem ser sobrepostos. Enquanto a *Gesinnung* pode ser considerada uma maneira de pensar (moral), o conceito de *Denkungsart* parece ser dotado também de um uso político, visto que Kant o mobiliza, por exemplo, em seu ensaio sobre o Esclarecimento (WA, AA 08: 36). Dessa maneira, é possível propor o *Denkungsart* como um termo guarda-chuva que abarca o conceito de *Gesinnung* mas adquire também, por vezes, uma significação política que o extrapola. Na literatura de comentário, Felicitas Munzel (1999) toma os dois termos como sinônimos, enquanto Gessis (2013) defende que *Gesinnung* e *Denkungsart* possuem diferenças significativas um em relação ao outro. Neste artigo, tendo em vista a possível sobreposição entre os termos, eles serão utilizados como sinônimos.

[5] Para uma visão geral desse problema, ver ALLISON 1990, p. 136 e ss.

Acontece que essa revolução na *Gesinnung* (ou ainda, no modo de pensar) não ocorre isoladamente: ela deve ser acompanhada por uma reforma nos costumes. É justamente a articulação entre esses dois elementos (um empírico e outro transcendental) que buscaremos analisar neste artigo. Para isso, tomaremos como ponto de partida a referida Observação Geral. A partir dos elementos textuais desse excerto, avançaremos a hipótese de que reforma e revolução consistem em dois pontos de vista a partir dos quais é possível referir-se a um mesmo acontecimento: o aprimoramento moral dos indivíduos. Veremos que Kant lança mão do artifício teórico de um "ponto de vista divino" para mostrar que aquilo que empiricamente se mostra como uma reforma lenta e gradual, quando visto em sua totalidade pode ser considerado (ou não, a depender do caso) uma revolução. Pretende-se defender ainda que da impossibilidade de conhecer a revolução da *Gesinnung* de um indivíduo é possível retirar, na verdade, um argumento pela valorização da formação de um caráter empírico.

Para cumprir essa tarefa, o presente artigo está estruturado da seguinte maneira. A seção 1 é dedicada à análise do conceito de *Gesinnung* e das dificuldades interpretativas que ele suscita. Procura-se, ao longo da seção, mostrar que a *Gesinnung* pode ser entendida como um *compromisso* assumido pelo agente e reafirmado por ele a cada ação. A seção 2 explora "reforma" e "revolução" como dois pontos de vista a partir dos quais o aprimoramento moral de um indivíduo pode ser avaliado, mostrando, a partir da análise da Observação Geral à Primeira Parte da *Religião*, que se trata de conceitos complementares e articulados entre si. A seção 3 parte dos resultados da seção anterior (a saber, a consideração de reforma e revolução como dois pontos de vista) para fazer uma leitura crítica da hipótese segundo a qual o aprimoramento moral dos indivíduos poderia ser pensado em estágios. Por fim, a seção 4 examina os desdobramentos que essa forma de entender a relação entre reforma e revolução no âmbito moral têm para a noção de caráter empírico. Acreditamos que essa leitura representa uma forma de enfatizar o papel que o caráter empírico desempenha na formação do sujeito moral kantiano.

1. Gesinnung entendida como compromisso do agente

Se tanto na *Fundamentação* quanto na *Crítica da razão prática* o foco de Kant está naquilo que nos permite julgar se determinada ação é boa ou má, na *Religião* ele parece enfatizar uma instância anterior de escolha, que nos permitiria discernir se determinada pessoa é boa ou má — instância a que ele denomina intenção (*Gesinnung*). Embora esse conceito adquira na Religião um lugar de destaque, não se trata de tomá-lo como uma exclusividade ou mesmo uma novidade deste texto, visto que a noção de que os seres humanos são dotados de uma intenção já se encontra em diversas passagens de seus escritos anteriores dedicados à fundamentação da moral[6], bem como desempenha um papel central alguns anos mais tarde, na *Doutrina da Virtude*. Kant pressupõe, portanto, que haveria uma instância que subjaz às escolhas individuais do sujeito e que constitui parte de seu caráter. A questão é compreender como ela se encaixa com outros aspectos de sua filosofia moral.

Para afirmar que uma determinada pessoa é boa ou má, deve-se pressupor em suas ações uma unidade ou, ao menos, certa coesão interna na escolha das máximas que as orientam. Para que tal coesão seja possível, projeta-se a ideia de que haveria um fundamento comum a todas as máximas

[6] A noção de *Gesinnung* aparece tanto na *Fundamentação da Metafísica dos Costumes* (GMS, AA 04: 406; 412; 416; 435) quanto na *Crítica da razão prática* (KpV, AA 05: 56; 75; 83-4; 99; 116; 147; 152-3), textos anteriores tanto à redação do ensaio sobre o mal radical (1792) quanto à publicação de *A Religião nos limites da simples razão* (1793).

de um agente, fundamento esse que consistiria em sua intenção (*Gesinnung*). Compreendida dessa maneira, a intenção representa a escolha fundamental de colocar seja a lei moral seja qualquer outro incentivo (aos quais Kant se refere, na *Religião*, pelo termo geral de amor de si) como prioridade no momento de agir. Nesse sentido, a intenção possuiria também a estrutura de uma máxima.

Tendo em vista que o ser humano necessariamente admite em suas máximas, de maneira natural, tanto o amor de si como a lei moral, Kant conclui que "a diferenciação, se o ser humano é bom ou mau, encontra-se não na diferenciação dos móbeis que ele admite em sua máxima (não na matéria da máxima), mas em sua *subordinação* (na forma da máxima): *em qual dos dois móbeis ele toma como condição do outro*" (RGV, AA 06: 36). Assim sendo, um agente com uma boa *Gesinnung* teria uma máxima fundamental que poderia ser formulada da seguinte maneira: "Farei o que é moralmente requerido e o que eu desejo, contanto que meus desejos não entrem em conflito com a lei moral.", enquanto a má *Gesinnung* poderia ser formulada com a ordem de prioridade invertida: "Farei o que eu desejo e o que é moralmente requerido contanto a lei moral não entre em conflito com meu amor de si"[7]. Como se trata de uma hierarquização entre dois fatores (o amor de si e a lei moral), não há a possibilidade de meio-termo — ou se é bom (dando prioridade à lei moral frente ao amor de si) ou se é mau (dando prioridade a qualquer outro móbil que não o respeito à lei moral)[8].

A análise da intenção ganha destaque na Primeira Parte da *Religião* justamente na ocasião em que Kant apresenta sua tese sobre a radicalidade do mal. Segundo ele, o mal é radical no sentido de que todas as ações más apresentam uma raiz comum, qual seja, a corrupção da *Gesinnung* dos agentes (RGV, AA 06: 37). A raiz comum a todas as ações más seria, nesse sentido, a priorização do amor de si frente à lei moral. Tal explicação está em consonância com a própria constatação de que a fonte do mal está no amor de si "adotado como princípio de todas as máximas" (RGV, AA 06: 45). O mal moral decorre, portanto, de colocar a satisfação do amor de si como condição para o cumprimento da lei moral.

Por mais que a ideia de que os seres humanos são dotados de uma *Gesinnung* esteja presente em outros textos de Kant dedicados moral, é na *Religião* que as dificuldades interpretativas relativas a esse conceito se tornam mais evidentes. Para que possamos pensar a própria possibilidade do aprimoramento moral dos indivíduos — tema sobre o qual nos debruçaremos nas seções seguintes —, é preciso enfrentar diretamente ao menos duas dessas dificuldades. A primeira diz respeito à própria concepção do que seria uma má *Gesinnung*, sobretudo frente à doutrina do mal radical. Dependendo de como se compreende o conceito de má *Gesinnung* e o grau de influência que uma determinada intenção exerce sobre a escolha das máximas individuais, mina-se qualquer possibilidade de alterar a *Gesinnung* de um sujeito. Como observa Lawrence Pasternack (2014, p. 117) em seu comentário à *Religião*, a fusão entre mal radical e má *Gesinnung* traz problemas para explicar o aprimoramento moral dos indivíduos, visto que o mal radical não é extirpável. A própria apresentação que Kant faz do conceito de mal radical já nos dá uma medida do impasse gerado pela coexistência entre, por um lado, a ideia de que os seres humanos são dotados de uma

[7] Tais formulações são propostas por Korsgaard (1996, p. 165) e retomadas por Louden (2010, p. 102).

[8] A explicação de Kant para que não haja meio termo é simples: "No entanto, ele também não pode ser moralmente bom em alguns aspectos e, ao mesmo tempo, mau em outros. Pois se ele é bom em um, então admitiu a lei moral em sua máxima; portanto, se ele tivesse de ser, ao mesmo tempo, mau em outro aspecto, então, uma vez que a lei moral do cumprimento do dever em geral é apenas uma e universal, a máxima referente a ela seria universal, mas, ao mesmo tempo, apenas uma máxima particular, o que se contradiz." (RGV, AA 06: 24).

propensão para o mal (que consiste precisamente em uma tendência a agir por outros incentivos que não a lei moral) e, por outro, a necessidade de preservar a liberdade do sujeito em escolher a cada momento suas máximas de ação, sejam elas boas ou más. Trata-se da seguinte passagem:

> Este mal é radical, porque corrompe o fundamento de todas as máximas; ao mesmo tempo, como propensão natural, não pode ser extirpado pelas forças humanas, uma vez que isto só poderia ser feito por meio de máximas boas, o que não pode acontecer se o fundamento subjetivo supremo de todas as máximas for pressuposto como corrompido; todavia deve ser possível predominar sobre tal propensão, uma vez que ela é encontrada no ser humano como ser que age livremente (RGV, AA 06: 37).

Não é à toa que a afirmação de que o mal não pode ser extirpado por forças humanas reconduz muitos intérpretes à noção de graça como assistência divina nesse processo de transformação dos indivíduos. Entretanto, o problema persiste: se o único caminho para uma ação moralmente boa decorresse da atuação da graça divina, isso resultaria em uma teoria moral incompatível com os textos de fundamentação nos quais a ênfase está precisamente na autonomia do sujeito. Mesmo na referência em que faz à doutrina da graça, Kant subscreve à tese de que é necessário *tornar-se digno* de tal assistência[9]. Por isso, é necessário um exame mais profundo da noção de *Gesinnung* e de como ela pode ser compreendida em sua função de *fundamento* (passível de corrupção) das máximas.

A concepção segundo a qual um fundamento corrompido barraria a possibilidade de boas máximas, quando lida em função da doutrina do mal radical, parece tornar impossível toda a ação moral. Tal dificuldade aparece justamente se pensamos a *Gesinnung* como algo que o sujeito escolhe apenas uma vez — ou pior, como algo que é dado por sua condição finita — e que, a partir de então, passa a compor um sistema de máximas no qual o fundamento das máximas (a *Gesinnung*) estaria na base de todas as outras máximas possivelmente escolhidas pelo agente, que seriam consideradas então como mera consequências ou ramificações de uma escolha primordial[10]. Em uma tal leitura, o sujeito ter uma má *Gesinnung* barraria toda a possibilidade de ações moralmente boas, bem como vice-versa.

Todavia, Kant concebe a noção de *Gesinnung* como a "conduta de vida" (RGV, AA 06: 71) do agente. Assim, não é possível entendê-la nem como algo que está dado nem como algo a que os outros (ou mesmo o sujeito da ação) têm fácil acesso. Pelo contrário, o único acesso que se tem à *Gesinnung* de alguém é a partir do efeito que ela tem em suas ações. E quando Kant se refere à conduta de vida, ele tem em mente a totalidade da vida de alguém e não uma fragmentação temporal. Assim, a *Gesinnung* de um agente deve ser compreendida nos termos de um compromisso que ele faz (e constantemente reitera) de uma ou outra hierarquização[11].

---

[9] Na Observação Geral à Primeira Parte da *Religião*, as ideias de "conversão moral" ou "renascimento", que se referem à revolução na *Gesinnung*, estão atreladas a uma reabsorção crítica do conceito teológico de *graça* (enquanto assistência divina). Entretanto, uma leitura que abarque também este aspecto foge ao escopo de nosso artigo, pois constituiria um artigo em si mesmo. Para uma visão das diferentes interpretações de como esse conceito aparece na *Religião*, bem como a maneira como ele repercute o debate entre Agostinho e Pelágio, é possível referir aos comentários de Jean Louis Burch (1968), Gordon Michalson (1990), Jacqueline Mariña (1997), Stephen Palmquist (2010) e Leslie Stevenson (2014). De outro modo, afastando-se de uma leitura que enfatiza apenas o conceito de graça no âmbito individual, Chris Surprenant (2006) coloca a *formação de uma comunidade* como um meio de "tornar-se digno da assistência divina".

[10] A *Gesinnung* é apresentada dessa maneira, por exemplo, por Matthew Caswell (2006, p. 196). Ele ilustra sua leitura através da imagem de uma árvore na qual as raízes representariam o fundamento de todas as máximas e dela sairiam o tronco e galhos correspondentes a máximas cada vez mais particularizadas de ação.

[11] A noção de *Gesinnung* como um *compromisso adotado pelo sujeito* é endossada por Julia Peters (2018, p. 505), interpretação próxima da que defendemos aqui. Em certo sentido essa tese é adotada também por Laura Papish (2018, p. 178 e ss.), que trata

É, portanto, apenas pelas ações do agente que se poderia tentar inferir sua *Gesinnung*. Mesmo assim, trata-se ainda de uma inferência muito incerta. De fato, de uma ação que claramente transgrida a lei moral, pode-se inferir uma má *Gesinnung*. Entretanto, há dois aspectos que devem ser levados em consideração antes de afirmar que haveria, por trás de uma ação que é conforme ao dever — que têm ao expectador a aparência de moralidade —, uma boa *Gesinnung*. O primeiro deles é o simples fato de que não há como ter acesso a se uma ação conforme ao dever é feita também por dever. Vale lembrar que Kant admite que há algo de inescrutável mesmo ao próprio sujeito em suas ações — nunca é possível dizer com certeza porque escolheu-se agir da maneira x ou da maneira y[12]. Já o segundo aspecto diz respeito à permanência no tempo da escolha necessária para que se caracterize uma boa *Gesinnung*. Mesmo que essa ação particular tenha sido feita por dever, não há como garantir que o respeito à lei moral é incondicional e que ela se manterá no curso das ações. Assim, enquanto a inferência de uma má *Gesinnung* é relativamente fácil, ao menos no caso de uma ação contrária à lei moral, a inferência de uma boa *Gesinnung* é algo que não se pode nunca fazer com certeza.

Para que o agente seja verdadeiramente livre, no entanto, é preciso que a cada ação — a cada escolha de máximas particulares — ele reafirme o *compromisso* que representa a escolha de um fundamento bom ou mau a suas máximas. Assim, passamos a entender a noção de *Gesinnung* como uma escolha que é reiterada a cada nova ação, e não como uma escolha que, uma vez feita, permanece imutável (como poder-se-ia pensar a partir de algumas afirmações de Kant). Até porque cada ação de um indivíduo "pode e deve sempre ser julgada como um uso *originário* de seu arbítrio" (RGV, AA 06: 41). Ler o conceito de *Gesinnung* dessa maneira nos auxilia a compatibilizá-lo com a noção de fraqueza da vontade (que é, inclusive, a forma mais branda de manifestação da propensão para o mal nos seres humanos). O fato do mal radical não poder ser extirpável não significa que não se possa "predominar sobre ele" (RGV, AA 06: 37).

A segunda dificuldade que se impõe à noção de uma *Gesinnung* quando se tem em vista a possibilidade de aprimoramento moral dos indivíduos é que, para Kant, a escolha da Gesinnung constituiria um *ato inteligível* (RGV, AA 06: 31; 39) e, portanto, seria uma escolha *fora do tempo*[13]. Tal dificuldade pode ser solucionada pelo duplo ponto de vista gerado pelas noções correlatas de caráter sensível e caráter inteligível, que veremos a seguir.

2. Revolução no modo de pensar, reforma no modo de sentir

O tema do aprimoramento moral dos indivíduos é especialmente presente na Observação Geral à Primeira Parte da *Religião*, onde Kant discute a possibilidade de restabelecer a predisposição (*Anlage*) originária para o bem presente em todos os seres humanos. A ideia de que o processo de tornar-se moralmente bom consiste, em verdade, no restabelecimento de uma predisposição

a revolução na *Gesinnung* em termos cognitivos, indicando que a decisão de seguir a lei moral é um comprometimento do sujeito. Há, entre as duas interpretações, nuances interpretativas que ficarão mais claras ao longo do artigo, sobretudo na seção 3.

[12] A noção de que há algo de desconhecido na motivação moral nos sujeitos é notável na seguinte passagem da *Religião*: "Mas, sem dúvida, o homem não pode chegar a convencer-se disto de modo natural, nem por consciência imediata nem mediante a prova de sua conduta de vida levada até então; pois a profundidade do coração (o fundamento primeiro subjetivo das suas máximas) é a ele inacessível; (...)" (RGV, AA 06: 51). Tal noção já figurava na *Fundamentação da Metafísica dos Costumes* (GMS, AA 04: 407) e voltará a aparecer na *Metafísica dos Costumes* (MS, AA 06: 392). Ao mesmo tempo, é curioso notar que Kant coloca como *primeiro comando* dos deveres para consigo mesmo o comando de autoconhecimento em relação aos deveres morais (MS, AA 06: 441).

[13] A tese de que a escolha da *Gesinnung* consiste em um ato inteligível é um ponto bastante controverso. Críticas a esse aspecto podem ser encontradas, por exemplo, nos trabalhos de Gordon Michalson (1990, p. 85) e Samuel Loncar (2013).

que é originária (RGV, AA 06: 46) — e, portanto, jamais totalmente corrompida (RGV, AA 06: 45) — é de extrema importância diante da problemática do mal radical a que nos referimos na seção anterior.

Kant observa que, se uma "árvore originariamente boa (segundo a predisposição) produziu frutos ruins" (RGV, AA 06: 45), então não é impossível que o inverso também ocorra. Isso significa sustentar que a transformação moral deve ser possível ao ser humano mesmo diante do mal radical como corrupção do fundamento de todas as máximas. Aliás, tal transformação consiste ela mesma em um dever. Por conta disso, ele argumenta que ela pode ser alcançada pelo próprio esforço humano, já que o dever "não nos ordena nada além daquilo que nos é factível" (RGV, AA 06: 47) — em outras palavras, *dever* implica *poder*. Não percamos de vista algo que já estava presente na formulação kantiana acerca da radicalidade do mal: embora ele não possa ser inteiramente eliminado, deve ser possível ao ser humano prevalecer sobre tal propensão.

Na referida Observação Geral, Kant trata do aprimoramento moral dos indivíduos por meio de dois caminhos: reforma e revolução. Se nos deixarmos levar pelo paralelo com a maneira como esses termos são aplicados à política, consideraremos opostos tais modos de transformação — a reforma como um processo lento e gradual, e a revolução como uma ruptura completa com o estado anterior. No que diz respeito à moral, a reforma estaria situada no campo dos costumes, ao passo que apenas a ideia de revolução permitiria conceber a conversão moral do sujeito, figurada também pela noção de um "renascimento" do mesmo.

De fato, a revolução diz respeito a uma alteração na hierarquia das máximas do agente, o que implica uma transformação em sua intenção (*Gesinnung*). Nesse processo, não há gradações nem a possibilidade de um meio-termo: a mudança da intenção representa uma ruptura com o modo de pensar anterior que aquele mesmo sujeito apresentava. Se antes ele colocava a satisfação de seu amor de si como prioritária à lei moral, agora a ordem é inversa: ele prioriza o cumprimento da lei moral em detrimento de sua satisfação pessoal.

Segundo Kant, mesmo a melhor das reformas nos costumes seria incapaz de promover uma transformação tão profunda que alcançasse o fundamento comum de todas as máximas do agente. Ainda assim, ele não desconsidera a importância da reforma dos costumes nem na *Religião* nem em outros momentos de sua obra. A questão, portanto, não se coloca como uma escolha entre dois processos independentes, mas como a compreensão de que, particularmente quando se trata de pensar a formação moral dos indivíduos, reforma e revolução são dimensões complementares, que não podem ser concebidas uma sem a outra.

A chave para compreender como Kant articula, nesse contexto, reforma e revolução está na afirmação de que "a revolução é necessária ao modo de pensar, mas a reforma gradual é necessária ao modo de sentir (que opõe obstáculos ao primeiro) e, por isso, também deve ser possível ao ser humano" (RGV, AA 06: 47-8). Aos poucos começa a se delinear a ideia de que reforma e revolução correspondem a formas de transformação vinculadas a dois níveis distintos da existência humana, que refletem, também no sujeito moral, a duplicidade entre noumeno e fenômeno. Com isso, é possível caracterizar com mais precisão o que seria, neste texto e neste contexto, uma revolução. Como bem observa Anna Wehofsits, o que Kant denomina "revolução" quando

fala da transformação moral do agente não é uma decisão isolada de agir moralmente, mas sim a decisão combinada com o esforço de torná-la efetiva[14].

A imagem da reforma é empregada para descrever uma mudança que diz respeito ao mundo fenomênico, resultante de um processo lento e gradual. Nesse âmbito, o vocabulário kantiano aponta, para além da ideia de transformação no modo de sentir, tanto para uma "mudança dos costumes" (RGV, AA 06: 48) quanto para a formação de um caráter empírico (RGV, AA 06: 47). A revolução, por sua vez, designa a transformação do modo de pensar (*Denkungsart*) — isto é, da *Gesinnung*. Trata-se da maneira pela qual Kant concebe a formação de um caráter inteligível, condição do tornar-se *moralmente* bom (RGV, AA 06: 47).

Um primeiro contato com o texto kantiano poderia dar a entender que reforma e revolução deveriam ocorrer em uma ordem temporal definida. Isso porque Kant afirma, em determinado momento da Observação Geral, que "a formação moral do ser humano não deve começar do melhoramento dos costumes, mas da transformação do modo de pensar e da fundação de um caráter" (RGV, AA 06: 48, destaque adicionado). Essa afirmação sugere que seria necessário iniciar pela revolução do modo de pensar e pelo estabelecimento de um caráter inteligível, para que os efeitos dessa mudança se manifestassem posteriormente nas ações particulares dos agentes. Se assim o fosse, a revolução precederia logicamente a reforma: a ação moral só seria possível se o agente priorizasse a lei moral, ou seja, após realizar a transformação de sua *Gesinnung*. Essa interpretação faria particularmente sentido se compreendêssemos a própria *Gesinnung* como dada antes da ação do sujeito, e não como o compromisso reiterado a cada ação ao qual nos referimos na seção anterior de nosso texto.

Todavia, as noções de reforma e revolução podem ser compreendidas de forma mais clara — e, consequentemente, melhor articuladas — não a partir de uma perspectiva temporal, mas considerando dois pontos de vista sobre o mesmo acontecimento: a transformação moral do indivíduo. Para explicar essa duplicidade, Kant recorre ao artifício teórico de um "ponto de vista divino", em contraste com o ponto de vista dos seres humanos finitos. Do ponto de vista humano, a transformação do indivíduo é percebida como um progresso gradual, ou seja, uma reforma. Já sob a perspectiva do "ponto de vista divino", essa mesma mudança é considerada uma revolução. A ideia subjacente a essa dupla consideração é que a percepção humana está condicionada por uma referência à temporalidade, enquanto o ponto de vista divino permite avaliar a unidade do percurso. Esse olhar global possibilita uma intuição da vida inteira do indivíduo, única forma de julgar sua possível conversão. Vale lembrar que Kant trata a *Gesinnung* como uma *conduta de vida*, o que torna ainda mais necessária a totalidade das ações para que ela possa ser julgada.

Assim, reforma e revolução correspondem a dois modos de considerar a mesma transformação: enquanto a primeira reflete a percepção humana limitada à temporalidade e ao mundo fenomênico, a segunda corresponde à apreciação inteligível e unitária de toda a trajetória moral do sujeito. A passagem a seguir, embora longa, ilustra de forma especialmente clara essa distinção:

[14] Ao comentar a possível articulação entre reforma e revolução Anna Wehofsits (2016, p. 114) defende que a revolução deve ser entendida não como uma mudança repentina, mas como uma *mudança qualitativa* da intenção do agente. Em concordância com o que apresentamos ao longo dessa seção, ela afirma que "não é a decisão em si que constitui a revolução, mas sim a associação dessa decisão com o esforço contínuo por uma forma de vida correspondente — ou seja, uma forma de vida cujas máximas sejam compatíveis com a lei moral. Sob uma perspectiva holística e inteligível, a decisão firme e o empenho consequente em realizá-la se apresentam como uma unidade, e é nessa unidade que podem ser compreendidos como revolução." (WEHOFSITS 2016, p. 115).

> Para aquele que perscruta o fundamento inteligível do coração (de todas as máximas do arbítrio); para quem, portanto, esta infinidade do progresso é unidade, isto é, para Deus, isto é tanto quanto ser de fato um ser humano bom (agradável a Ele); e, nessa medida, essa mudança pode ser considerada uma revolução, embora, no julgamento do ser humano que pode estimar a si e a força de suas máximas apenas segundo o predomínio que elas ganham sobre a sensibilidade no tempo, essa mudança seja vista apenas como um esforço sempre contínuo para o melhor; por conseguinte, como reforma gradual da propensão para o mal como modo pervertido de pensar (RGV, AA 06: 48).

A revolução na *Gesinnung* do indivíduo — considerada enquanto ato inteligível — permanece inacessível à percepção humana temporal. Para nós, apenas o caráter empírico é observável, de modo que toda transformação moral aparece sempre como uma reforma lenta e gradual dos costumes. O que podemos perceber enquanto seres humanos é o caráter empírico, que nos revela apenas se determinada ação está em conformidade com o dever. No entanto, isso revela muito pouco sobre o caráter inteligível do agente: um indivíduo pode agir consistentemente de acordo com a lei, mas isso significaria apenas que "o caráter empírico é bom, mas o caráter inteligível continua sendo mau" (RGV, AA 06: 37). Em outras palavras, é quase impossível afirmar com certeza que alguém realizou a revolução; podemos apenas identificar quando ela não ocorreu, por meio de ações que transgridem diretamente a lei moral[15].

3. É possível falar em estágios para o aprimoramento moral?

Como vimos, a letra do texto kantiano comporta certa ambiguidade, ao menos à primeira vista, com relação à articulação possível entre reforma e revolução. Por conta das formulações que dão a entender que ambos os processos aconteceriam em certa ordem temporal, é possível encontrar na literatura de comentário aqueles que defendem que haveria, na filosofia de Kant, uma teoria de dois estágios para a transformação moral dos indivíduos. É o caso, por exemplo, de Laura Papish (2018) e Conrad Damstra (2023). Analisemos agora o que isso significa e a pertinência de tal hipótese interpretativa, tomando como exemplo o argumento de Papish.

No Capítulo 7 de seu livro *Kant on evil, self-deception, and moral reform*, Laura Papish defende que haveria uma teoria de dois estágios para o aprimoramento moral[16] de um indivíduo e os explica da seguinte maneira: "um primeiro estágio de conversão moral em que o respeito pela lei, por si só, é incorporado à máxima do indivíduo, e um segundo estágio de progresso moral em que o agente observa seu comportamento e ações em busca de evidências de sua nova intenção [*Gesinnung*]" (PAPISH 2018, p. 177). Ela propõe que entendamos a ideia de conversão moral como um comprometimento do sujeito que, para acontecer, deve ser precedido por duas tarefas, sendo uma delas a manifestação consistente de um comportamento conforme ao dever, mesmo que não por dever[17]. Para ela, o compromisso moral poderia ser explicado a partir do exemplo

[15] A duplicidade de ponto de vista aparece também na Segunda Parte da *Religião*: "Segundo a nossa avaliação, o ato — enquanto um progresso continuo ao infinito do bem para o melhor, para o qual estamos restritos inevitavelmente as condições de tempo nos conceitos da relação da causa e dos efeitos — permanece sempre defeituoso, de modo que temos de considerar o bem no fenômeno — isto é, segundo o *ato* — *sempre* como insuficiente em nós para um lei santa; mas seu progresso ao infinito em direção a adequação com essa lei devido a *intenção* [*Gesinnung*] a partir do qual ele é derivado, que é suprassensível, nós podemos pensar como sendo julgado — enquanto um todo completo também segundo o ato (a conduta de vida) — por um perscrutador de corações em sua intuição intelectual pura e, assim, o ser humano, a despeito de sua defectibilidade constante, pode esperar, ao fim, ser, *em geral*, agradável a Deus, não importa em qual momento do tempo sua existência seja interrompida." (RGV, AA 06: 67, tradução modificada).

[16] Em seu livro, ela se refere a este movimento como "reforma moral" (*moral reform*), mas aqui optamos por chamar de "aprimoramento moral" (por vezes "transformação moral") para marcar a diferença entre este aspecto mais amplo e o uso específico que Kant faz da metáfora de uma "reforma dos costumes", relacionada ao caráter empírico.

[17] Cf. PAPISH 2018, p. 190.

do compromisso de casamento, dentro de um recorte muito preciso do que isso significaria: um compromisso voluntário entre pessoas que tem "uma compreensão cognitiva boa, ainda que imperfeita, do trabalho que um casamento bem-sucedido exige" (PAPISH 2018, p. 190).

A partir da ideia de que a conversão vem de um compromisso do sujeito a determinado princípio de ação, Papish aponta para um caminho interessante ao notar que tal comprometimento deve ser precedido por uma série de esforços que não são partes do comprometimento ele mesmo e que, portanto, a conversão do agente não seria uma total surpresa nem para ele nem para aqueles à sua volta. Ela concebe, portanto, que haveria um estágio "pré-conversão", no qual o agente se familiariza com a moral. Nesse sentido, ela explora a noção de compromisso mostrando que para que o agente se comprometa com determinada posição que aponta para o cumprimento da moral, ele tem de ter um conhecimento prévio do que é a virtude. Todavia ao afirmar que a conversão moral deve ser precedida de um comprometimento que depende de uma série de esforços que não são parte do comprometimento em si, é quase como se se falasse em três e não dois estágios: um estágio pré-conversão, a conversão e um momento posterior em que se observa os efeitos de tal conversão (entendidos por ela como um trabalho cognitivo e não volitivo)[18].

A hipótese de uma teoria de estágios enfrenta, no entanto, algumas dificuldades significativas. A primeira delas é justamente a de como temporalizar o ato inteligível que caracteriza a escolha de uma boa *Gesinnung*. A noção de estágios traz consigo a ideia de que seria possível demarcar temporalmente o momento em que ocorre tal transformação, o que *de fato* não é possível. A revolução moral para Kant não é o ato isolado da decisão, mas a soma deste com o esforço de torná-lo efetivo. Em segundo lugar, a ideia de estágios parece colocar problemas para se pensar a fraqueza da vontade. Seria possível haver, então, várias revoluções na vida de um mesmo indivíduo? Se assim o for, esbarramos em uma terceira dificuldade, a saber, a de conciliar a fragmentação temporal da possibilidade de várias revoluções com a ideia de que a *Gesinnung* é, ao fim e ao cabo, a avaliação da conduta de vida (de *toda* a vida) de um agente.

Uma alternativa a esta interpretação, pela qual pretende-se aqui argumentar, seria manter a ideia de que se precisa conhecer a virtude para se comprometer com ela sem, no entanto, pensar que a partir dela viriam revolução e reforma como estágios posteriores. Nesse sentido, é da educação moral de um indivíduo e da formação de seu caráter empírico que surge uma reforma de seus costumes que, talvez, quando olhada na totalidade de suas ações, possa representar que ali ocorreu uma revolução em sua *Gesinnung*.

4. Uma aposta no caráter empírico

O vocabulário temporal utilizado por Kant não significa, como mostramos na seção anterior, que haja *de fato* uma antecedência da revolução da intenção do agente em relação à reforma de seus costumes, de modo que seja possível ao agente utilizar determinado momento como um marco em sua história pessoal. O fato de haver algo a que Kant se refere como uma revolução na *Gesinnung* não significa, curiosamente, que se possa falar em um antes e um depois dessa revolução. Pelo contrário, o que se chama aqui de *revolução na Gesinnung* parece ser apenas a contrapar-

[18] Esse parece ser o maior ponto de discordância da parte de Conrad Damstra em relação à tese defendida Laura Papish. Para ele, não se trata de um trabalho cognitivo *ao invés de* volitivo — ele escreve: "Na verdade, creio que a discussão de Kant sobre a virtude resiste a tal dualismo." (DAMSTRA 2023, p. 572, nota 21).

te noumênica do aprimoramento moral dos indivíduos que se apresenta *para nós* como uma reforma lenta e gradual dos costumes. A própria ideia de que a revolução na intenção dos agentes constituiria um ato inteligível já aponta para a dificuldade de colocá-la em uma série temporal.

Além de representar uma "escolha fora do tempo", a revolução na intenção deve se referir não a curtos períodos de tempo, mas à totalidade da vida de um agente. Assim, um olhar retroativo do indivíduo à sua trajetória também não daria conta de afirmar, com certeza, se ali houve ou não uma transformação qualitativa de sua *Gesinnung* e se essa transformação possuiria a duração no tempo necessária para caracterizar a conversão moral. Ora, se a revolução não é acessível aos seres humanos finitos (nem mesmo ao próprio agente), tudo o que sobra como parâmetro do progresso moral são as ações individuais de cada um. Se só temos acesso ao caráter empírico, a formação de um caráter empírico adquire um lugar central quando se pensa na possibilidade de criar agentes morais virtuosos.

A afirmação a que nos referimos na seção 2, segundo a qual "a revolução é necessária ao modo de pensar, mas a reforma gradual é necessária ao modo de sentir" (RGV, AA 06: 47) trazia entre parênteses um adendo que nos ajuda a compreender a importância de olharmos para o caráter empírico: *o modo de sentir opõe obstáculos ao modo de pensar*. Se o modo de sentir opõe obstáculos ao modo de pensar, uma transformação no modo de sentir se faz fundamental para que a mudança para um modo de pensar moral possa ocorrer e manter-se como tal.

Ao longo de toda a Observação Geral da Primeira Parte da Religião, Kant assinala a importância dos *exemplos*, da noção de *cultivo moral* e de se *ensinar a moralidade* (referindo-se aos "aprendizes morais"). Com efeito, o caráter empírico pode ser ensinado, tornando-se um instrumento para a promoção da moralidade:

> A firme resolução, transformada em prontidão, no cumprimento de seu dever também se chama virtude, segundo a legalidade, em seu *caráter empírico* (*virtus phaenomenon*). Ela tem a máxima persistente de ações *em conformidade* com a lei, não importa de onde se toma o móbil que o arbítrio precisa para isso. Por isso, a virtude, nesse sentido, é adquirida *pouco a pouco* e significa, para alguns, um longo hábito (na observância da lei), por meio do qual o ser humano passou, mediante reformas graduais de sua conduta e da consolidação de suas máximas, da propensão para o mal para uma propensão oposta. Ora, para isso não é necessário exatamente uma *mudança do coração*, mas apenas uma mudança dos *costumes* (RGV, AA 06: 47).

Se o progresso moral depende da revolução na *Gesinnung* dos agentes e o único acesso que temos a essa mudança é por meio das ações dos indivíduos, então o caminho (ao menos do ponto de vista empírico) para o aprimoramento moral passa por um ajuste das próprias ações. Por mais que a formação de um caráter empírico garanta apenas a legalidade das ações e não sua moralidade, é através desse processo que a moral é ensinada — isso mesmo quando se pensa que o ensinamento da moral é diferente da criação de um hábito de praticar ações conformes ao dever[19]. Embora Kant faça um uso muito restrito do conceito de virtude na *Religião*, relacionando-o apenas com a legalidade das ações (RGV, AA 06: 47), outros textos apresentam um conceito mais amplo e propriamente ético de virtude. É o caso da *Metafísica dos Costumes*, em que ele define a virtude como "a faculdade e o propósito refletido de opor resistência (...) ao adversário da intenção moral *em nós*" (MS, AA 06: 380). Adquirir a virtude, portanto, é reiterar o compromisso com uma boa intenção.

[19] Essa diferenciação é clara na *Metafísica dos Costumes*: "Portanto, não se pode *definir* a virtude como o hábito de praticar ações livres conformes à lei; a menos que se acrescentasse "de determinar-se a agir por meio da representação da lei"; e neste caso esse hábito é uma propriedade não do arbítrio, mas antes da *vontade*, que, com a regra que ela adota, é uma faculdade de apetição ao mesmo tempo universalmente legisladora. Apenas um tal hábito pode ser contado como virtude" (MS, AA 06: 407).

Para além disso, Kant indica que a própria revolução também é, no limite, um resultado deste processo de cultivo que começa com ações conformes ao dever. Em suas palavras: "... esta predisposição passa, pouco a pouco, ao modo de pensar, de maneira tal que o dever começa a adquirir, meramente por si mesmo, um peso notável em seus corações." (RGV, AA 06: 48). Assim, por mais que a admiração pelas ações virtuosas não seja ainda a boa intenção, ela parece ser um passo importante para que o sujeito escolha, ele mesmo, fazer o compromisso de colocar a lei moral acima da satisfação de seu amor de si. Nesse sentido, é como se Kant apontasse para a importância do processo de interiorização de determinada conduta. Ele ressalta que cada escolha pela lei moral dá forças para que o indivíduo escolha novamente a lei moral. A seguinte passagem ilustra esse aspecto

> Pois o ser humano que, desde a época que adotou os princípios do bem, percebeu, durante uma vida suficientemente longa, os efeitos destes princípios sobre o ato – isto é, sobre sua conduta de vida que sempre progride para o melhor – e que encontra ocasião para inferir disso, mesmo que supostamente, um melhoramento fundamental em sua intenção [*Gesinnung*], pode, contudo, também esperar racionalmente que – visto que os mesmos progressos, desde que seu princípio seja bom, aumentam sempre a *força* para os progressos seguintes – ele não mais abandonará esse caminho nessa vida terrena, mas sempre progredirá corajosamente nele; e, por certo, se outra vida ainda lhe for iminente depois desta, ele continuará a progredir doravante, sob outras circunstâncias e, segundo toda aparência, nesse caminho, precisamente de acordo com o mesmo princípio, e se aproximara sempre mais da – embora inatingível – meta da perfeição, uma vez que, segundo o que percebeu em si até então, pode considerar sua intenção [*Gesinnung*] melhorada a partir do fundamento (RGV, AA 06: 68, tradução modificada).

Com esses elementos em mente, é possível pensar que a reforma que se inicia pelos costumes contribui para que se forme no agente uma boa intenção — sobretudo quando se compreende a intenção da maneira como a definimos: como um compromisso reiterado do agente. Como Kant já apontava em seu ensaio sobre a Ideia de uma história universal, é a partir do processo de Esclarecimento (que envolve o refinamento da cultura e o cultivo do gosto) que se pode chegar a um modo de pensar moral. Mais do que uma ruptura instantânea, é a esse movimento — que chamamos de formação de um caráter empírico — que podemos atribuir a mudança qualitativa necessária para que se projete que ocorreu uma revolução na intenção de determinado agente.

Considerações finais

Ao longo deste artigo, procuramos extrair, de uma análise da Observação Geral à Primeira Parte da *Religião*, um argumento que reforça a importância da formação de um caráter empírico na filosofia moral de Kant. Para isso, partimos de uma análise de um tema central da discussão kantiana acerca do mal radical: o conceito de *Gesinnung* e sua possibilidade de mudança. Por outro lado, procuramos pensar a mudança no modo de pensar do agente e a reforma gradual dos costumes como duas faces da mesma moeda. Trata-se, portanto, do mesmo acontecimento (o aprimoramento moral do indivíduo) considerado sob dois pontos de vista. Levando em conta que (1) a *Gesinnung* deve ser entendida como um compromisso reiterado do agente e (2) a mudança na *Gesinnung* se apresenta como um processo lento e gradual do ponto de vista finito, torna-se mais fácil perceber a importância de se atentar para a formação do caráter empírico do sujeito. Curiosamente, o argumento também se sustenta quando se considera a função que a graça divina desempenha nesse texto. Mesmo para as leituras que a enfatizam, é possível apontar que a importância do caráter empírico reside no fato de que concepção kantiana de graça depende da ideia de um tornar-se digno da assistência divina — algo que é alcançado também por meio de uma reforma nas ações do sujeito.

Referências bibliográficas

ALLISON, H. 1990. *Kant's theory of freedom*. Cambridge: Cambridge University Press.

BRUCH, J. L. 1968. *La Philosophie Religiouse de Kant*. Aubier: Editions Montaigne.

CASWELL, M. 2006. "Kant's Conception of the Highest Good, the *Gesinnung*, and the Theory of Radical Evil". In: *Kant-Studien* 2006, pp. 184-209.

DAMSTRA, C. 2023. 'The Change of Heart, Moral Character and Moral Reform', *Kantian Review*, 28(4), pp. 555–574.

DICENSO, J. 2012. *Kant's Religion within the Boundaries of Mere Reason: A Commentary.* Cambridge: Cambridge University Press.

GRESSIS, R. A. 2013. "The Relationship Between the *Gesinnung* and the *Denkungsart*", In: *Kant und die Philosophie in weltbürgerlicher Absicht: Akten des XI. Kant-Kongresses 2010*, Berlin, Boston: De Gruyter, 2013.

KANT, I. *Gesammelte Schriften herausgegeben von der Deutschen Akademie der Wissenschaften*, anteriormente *Königlichen Preussischen Akademie der Wissenschaften*, 29 vols. Berlim, Alemanha: Walter de Gruyter, 1902ss.

KANT, I. *Ideia de uma história universal de um ponto de vista cosmopolita* (R. Terra, trad.). São Paulo, Brasil: Martins Fontes, 2004.

KANT, I. *Fundamentação da Metafísica dos Costumes* (G. A. de Almeida, trad.). São Paulo: Discurso Editorial/Barcarolla, 2009.

KANT, I. *Metafísica dos Costumes*. Petrópolis, RJ: Vozes; Bragança Paulista, SP: Editora Universitária São Francisco, 2013.

KANT, I. *Crítica da razão prática* (M. Hulshof, trad.). Petrópolis, RJ: Vozes; Bragança Paulista, SP: Editora Universitária São Francisco, 2016.

KANT, I. *A religião nos limites da simples razão* (B. Cunha, trad.). Petrópolis: Vozes, 2024.

KORSGAARD, C. 1996. *Creating the Kingdom of Ends*. New York: Cambridge University Press.

LONCAR, S. 2013. 'Converting the Kantian Self: Radical Evil, Agency, and Conversion in Kant's *Religion within the Boundaries of Mere Reason*', In: *Kant-Studien* 2013; 104(3): 346–366.

LOUDEN, R. 2010. "Evil Everywhere. The Ordinariness of Kantian Radical Evil", In: *Kant's Anatomy of Evil*. Cambridge: Cambridge University Press. pp. 93-115.

MARIÑA, J. 1997. "Kant on Grace: A Reply to His Critics", In: *Religious Studies*, Vol. 33, No 4 (Dec., 1997), pp. 379-400.

MICHALSON, G. 1990. *Fallen Freedom: Kant on Radical Evil and Moral Regeneration.* Cambridge: Cambridge University Press.

MUCHNIK, P. 2009. *Kant's theory of evil: an essay on the dangers of self-love and the aprioricity of history.* Lanham: Lexington Books.

MUNZEL, F. 1999. *Kant's conception of moral character: the "critical" link of morality, anthropology, and reflective judgment*. University of Chicago Press.

PALMQUIST, S. 2010. "Kant's Ethics of Grace: Perspectival Solutions to the Moral Difficulties with Divine Assistance", In: *The Journal of Religion*, 90(4), pp. 530-553.

PAPISH, L. 2018. *Kant on evil, self-deception, and moral reform*. Oxford University Press.

PASTERNACK, L. 2014. *Routledge Philosophy Guidebook to Kant on Religion within the Boundaries of Mere Reason*. Londres e Nova York: Routledge.

PETERS, J. 2018. "Kant's *Gesinnung*", In: *Journal of the History of Philosophy, 56(3)*, pp. 497-518.

STEVENSON, L. 2014. 'Kant on grace', in G. Michalson (ed.) *Kant's Religion within the Boundaries of Mere Reason: A Critical Guide*. Cambridge: Cambridge University Press (Cambridge Critical Guides), pp. 118–136.

SURPRENANT, C.W. 2006. 'Cultivating Virtue: Moral Progress and the Kantian State', *Kantian Review*, 12(1), pp. 90–112.

WEHOFSITS, A. 2016. *Anthropologie und Moral.* Berlin, Boston: De Gruyter.

WOOD, A. 2020. *Kant and Religion*. New York: Cambridge University Press.

Evil, yet righteous: Kant's devils and the moral concept of right

Maus, mas ainda assim justos: os demônios de Kant e o conceito moral de direito

Gehad Marcon Bark[1]
Universidade Federal do Paraná (UFPR)
gehad_marcon_bark@hotmail.com

Abstract: This paper focuses on *Kant's Toward Perpetual Peace* famous statement according to which the problem of the State can be solved even for a race of devils. Its first aim is to show that coercion exerted via positive laws is pivotal to the understanding of Kant's main thesis regarding this nation of devils and the correspondent accomplishment of reason's ends through the "mechanical course of nature" (ZeF, AA 08, p. 367). According to this reviewed version of Kant's hypothesis, however, by refraining from violating the laws of a republican State out of self-interest, a devil would have to be taken as genuinely righteous according to rational principles and, more particularly, the moral concept of right. To make sense of this statement, Kant's hypothesis shall be developed and interpreted as coherently enclosing a general thesis regarding the normativity of right and its source on the external use of free choice independently of each agent's moral virtue.

Keywords: coercion; devils; Kant; perpetual peace; reason; right.

Resumo: O artigo enfoca a famosa afirmação de Kant em *À paz perpétua* segundo a qual o problema do Estado pode ser solucionado mesmo para uma raça de demônios. Seu primeiro propósito é demonstrar que a coerção exercida por meio de leis positivadas é fundamental para a compreensão da tese central de Kant sobre essa nação de demônios e a correspondente realização dos fins da razão através do "curso mecânico da natureza" (ZeF, AA 08, p. 367). Segundo essa versão revista da hipótese de Kant, contudo, ao abster-se de violar as leis de um Estado republicano por autointeresse, um demônio teria de ser tomado como genuinamente justo segundo princípios racionais e, mais particularmente, segundo o conceito moral de direito. Para entender essa afirmação, a hipótese de Kant deve ser desenvolvida e interpretada como compreendendo coerentemente uma tese geral acerca da normatividade do direito e sua fonte no uso externo da liberdade do arbítrio independentemente da virtude moral de cada agente.

Palavras-chave: coerção; demônios; Kant; paz perpétua; razão; direito.

[1] This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 (88887.100967/2024-00). Project CAPES/DAAD/PROBRAL (88887.627936/2021-00): "Perspectivas Kantianas sobre as causas da irracionalidade social". I also would like to thank the members of the research project for the enriching discussions during the workshops held at the University of Vechta in 2025. They greatly contributed to the final version of this paper.

Recebido em 30 de julho de 2025. Aceito em 11 de dezembro de 2025.

1. Introduction

Kant's affirmation regarding the possibility of a solution to the problem of State even for a race of devils (ZeF, AA 08, p. 366)[2] is widely known and discussed by scholars. From an exegetical point of view, the reason for this lasting interest still lies in the fact that, if Kant is really stating, in *Toward Perpetual Peace*, that even devils (understood as beings whose actions are done exclusively out of self-interest) could stablish a State, then he would be apparently endorsing the thesis according to which the source of the normativity of right and juridical laws stems from what Kersting calls the "prudence in the service of self-interest" (KERSTING, 1992, p. 342).

Just like many other authors, Kersting promptly (and correctly) rejects this kind of reading of Kant's political and juridical thought. To make the general point of this rejection clearer, let's take Allen Wood's enlightening remarks regarding this particular passage as an example. According to him, Kant's statement concerning the race of devils implies simply that "reason requires that there be a just society in which external coercion would be sufficient to protect the rights even of rational beings who are entirely lacking in moral virtue" (WOOD, 1999, p. 323). Dealing with the distinction between right and ethics (which lies in the core of the *Metaphysics of Morals*) a couple years later, Wood once again points out to what would be typical of a community in Kant's rational framework: "a civil society based on right requires no commitment on the part of its members to respect one another's rightful freedom" (WOOD, 2002, p. 8). The juridical (and rational) problem of State, he says, amounts to nothing more than "a system of external legislation, backed by coercive sanctions sufficient to guarantee that rights will not be infringed" (WOOD, 2002, p. 8).

Wood's reading offers a clear example of the most usual approach to the seemingly problematic case of a race of devils. According to such interpretation, Kant raises this hypothesis to emphasize only that juridical laws, subject to an external lawgiving, can be obeyed under any "motivating mechanism[s]" (LUDWIG, 2002, p. 162), self-interest obviously included among those. As far as the hypothesis goes, moreover, "different motivational incentives with respect to juridical lawgiving are restricted to the level of law enforcement and do not support claims about self-interest as the normative ground of justice" (FLIKSCHUH, 2000, p. 94).

The last statement is undoubtedly correct as a reconstruction of Kant's thoughts. Due to well-known philosophical and systematic reasons, self-interest is by no means the source of juridical normativity in the stronger moral sense developed by Kant[3]. But shall we also conclude, on the basis of this accurate exegetical approach, that the hypothesis of a race of devils bears no interest when it comes to better understanding the normativity of right on a Kantian view?

[2] Direct and indirect quotations of Kant's philosophical work will follow the *Akademie-Ausgabe's* standard (acronym, volume, page). For a better understanding of acronyms and volumes used to refer to each of Kant's original texts in German language, see the Bibliographical References.

[3] For an example of a discussion on different types of normativity opposed to Kant's specifically moral normativity, consider how Christine Korsgaard's addresses the problem of the *ethics* of a mafioso or gangster in *The Sources of Normativity*. Briefly, in her response to the objections raised by to G. A. Cohen's in *Reason, humanity and the moral law*, Korsgaard proposes that even a member of a criminal group is committed to some kind of normativity when acting in accordance with a code imposed by his group. And this occurs because this agent, as a rational being who can reflect upon his acts, should give his "reflexive endorsement" (KORSGAARD, 1996, p. 257) to such imperatives. The main point for her (one that is developed at much more length in the lessons that integrate the book) is that the normativity of this criminal agent (that we could treat maybe as a prudential one) does not satisfy Kant's moral requirements (which involves, most notably, humanity as a value in itself).

Although none of the authors above explicitly advocates that no normative thesis could possibly be advanced on the basis of this hypothesis, it is interesting to note how quickly the subject is dismissed as soon as brought up to discussion. Self-interest – as the only selfish reason on which devils would act – is commonly linked to the specificity of external lawgiving and, correspondently, to Kant's introductory remarks on legality as a conformity with laws through incentives drawn from "pathological determining grounds of choice" (MS, AA 06, p. 218). No further developments beyond this well-established interpretation are usually attempted.

Wouldn't there be a more fruitful use of the case of the devils, namely, one able to show its relevance to the comprehension of the limits within which Kant circumscribes his views on right and law from a rational point of view? In this paper I try to show that this is precisely the case. I shall argue that Kant's discussion about this nation of devils is strictly aligned to the moral concept of right presented and developed in the *Doctrine of Right*. I argue, more directly, that *Toward Perpetual Peace's* hypothesis of a race of devils is not raised in order to discuss whether such egoistical beings would be capable of leaving the state of nature out of their prudential reasoning alone (that being the rational and normative ground for juridical laws). Instead, it shall be interpreted as a thesis concerning specifically the instrumental role of coercive positive laws to promote reason's ends, through inclinations, from a juridical point of view. In this manner, however, as paradoxical as it may seem, Kant's discussion about a nation of devils also encloses a more general thesis about the normativity of right itself.

2. The Kantian problem of State for a race of devils reconsidered

The departing point of our discussion shall be, naturally, the description of such devils provided by Kant in *Toward Perpetual Peace.* And what precisely is said about them? A couple of interesting things, actually. First of all, Kant says that these devils should be taken "as a group" (ZeF, AA 08, p. 366). Plainly speaking, this means that they are social beings and, consequently, do not live in isolation. In other words, they interact with each other.

Secondly, presumably because they interact with each other, they also need what Kant describes as "universal laws for their preservation" (ZeF, AA 08, p. 366). Two things can be said about this point. On one side, if preservation is a serious matter for these devils, then they are not omnipotent, nor infinite beings. This means that their finite existence is not only subject to elimination, but highly dependent on the fulfilment of a set of specific conditions. One would not need to necessarily assume that they are identical, in nature, to human beings, but let's admit, for the sake of the argument, that they have some analogous sensuous nature and, correspondently, a set of needs attached to their preservation. Now, another thing to be said about these devils is that, for such finite and social beings, interaction is supposed to pose problems (otherwise, Kant's whole discussion about laws for preservation for social beings would not make sense at all). Again, if preservation is an actual matter, then, in this hypothetical scenario, these devils are to be taken as reciprocally vulnerable to each other (regardless of which may be the intentions of each one of them). On a minimal level, these devils' preservation presupposes a sensuous nature at least in the sense according to which they have needs related to the mitigation (if not elimination) of risks that emanate from reciprocal interactions.

Thirdly, Kant adds to the description of these devils both that they "require universal laws for their preservation" (ZeF, AA 08, p. 366) and that, despite this requirement, each one of them is

"secretly inclined to make an exception of himself" (ZeF, AA 08, p. 366). How are we supposed to understand this last statement? For all those who are familiar with the main discussions themed in the *Groundwork to the Metaphysics of Morals*, Kant famously refers the case of someone *who makes an exception of himself* when arguing that, even by acting against the moral law, a rational being simply allows few exceptions on his behalf (thus acting against the supreme moral principle), but still recognizes the "validity of the categorical imperative" (GMS, AA 04, p. 425). Leaving aside the deeper issues raised by this controversial passage (especially those related to the well-known problem of self-deception), one way to understand Kant's affirmation is by saying that, as universal, these laws are clearly supposed to assure preservation for each member of the group. However, if each devil is also inclined to make an exception of himself, then, from the standpoint of each member of this group, those laws for preservation would not hold universally. In other words, by making himself an exception, each devil is also inclined to infringe upon each other's preservation.

But why would they do that (especially if they are rational beings)? Why would they prefer a condition other than the one in which their own preservation is assured under universal laws (by means of some kind of reciprocal limitation capable of securing at least a minimal level of satisfaction for each one)? Correctly understood, this particular depiction of a devil seems to presuppose more than the fact that each member of this hypothetical group has in mind only his self-preservation. In fact, the very last thing that Kant says about these devils is that they have "evil intentions" [*bösen Gesinnungen*] (ZeF, AA 08, p. 366). It seems that a devil is *someone who makes an exception of himself in the very particular sense of being evil* or, following Kant's own phrasing, possessing evil intentions[4].

Now, as well known, in his discussion about the sources of evil in *Religion within the Boundaries of Mere Reason*, apart from analyzing the degrees of evil, Kant says that, on its most general level, evil arises from a common source, i.e., it arrives precisely when someone incorporates "incentives of his sensuous nature" (RGV, AA 06, p. 36) against the moral law. If we put all the pieces together, a more radical but still plausible way to read Kant's description of a devil in *Towards Perpetual Peace* is to say that, for these beings, preservation becomes a matter precisely because each one of them is vulnerable but also solely driven (at the motivational level) by the maximal satisfaction of sensuous needs and interests of inclination. In this sense, each devil is willing to invest against other members of his social group in the name of his own benefit.

More specifically, devils are to be taken not only as beings driven by self-interest alone, but as rational but evil beings who systematically and without exceptions act out self-interest to maximally satisfy their own particular inclination-based interests even at the expense of the freedom and

[4] One of the main points of discussion about the interpretation advanced in this paper can be put as follows: what would be, after all, the difference between Kant's devils and human beings? Crudely taken, the passage of *Toward Perpetual Peace* seems to leave no room for any relevant difference to be outlined. However, careful attention to Kant's discussion on public right and the duty to enter a "civil condition" (MS, AA 06, p. 312), for instance, brings to light at least one important distinction. As Reinhard Brandt correctly observed, Kant never assumes that human beings are evil (BRANDT, 2012, p. 189), let alone discusses the duty to enter a civil condition under anthropological and empirical presuppositions of this sort. As a matter of fact, according to him, even those "well-disposed and law-abiding" (MS, AA 06, p. 312) would still need a State equipped with precise and public adjudicatory rules set specifically to solve conflicts independently of unilateral conceptions of right. At most, Kant suggests in the *Doctrine of Right* that human beings could be presumably evil – "praesumitor malus" (MS, AA 06, p. 307). But that statement should be taken in a much weaker sense according to which humans simply pose a potential threat to each other, regardless of their own intentions, by simply coexisting in the same planet *qua* finite beings, as Pauline Kleingeld also correctly pointed out (KLEINGELD, 2004, p. 318). This assumption, however, is not tantamount to being actually evil, which seems to be the case of devils in *Towards Perpetual Peace.*

existence of others. In this very specific sense, one can say that these devils are radically egoistical beings who, far from being driven by a rational interest in some kind of balance to assure mutual safety (which could lead them, on a Hobbesian fashion, to cooperate and to act, out of self-interest, to assure their own preservation), are willing to subdue, slave and kill each other only to maximize their personal benefits[5].

Once these devils are defined as evil in this particular manner, we need to consider more closely the context in which they are discussed, i.e., the context of social interactions related to the well-known problem of State in *Toward Perpetual Peace*. But to what amount, after all, said problem? Is Kant really addressing the duty to leave the state of nature as a problem and speculating that these devils, despite inherently evil in the sense above defined, would be able to perform it out of prudential reasoning alone?

If we are to correctly grasp the problem that Kant has in mind (and also the solution intended to it), the idea of an instrumental use of inclinations to accomplish reason's juridical prescription (which clearly structures the *First Supplement*) shall be correctly understood. It obviously cannot simply mean that, on its own, self-interest (especially in the case of radically egoistical beings) would spontaneously lead to a condition within the limits of which rights would be preserved in accordance with universal laws. If that were the case, then organizing these creatures wouldn't even pose any problems at all. But, as we know, Kant was not naive, nor utopian, to be led to believe that, outside the domain of positive laws and coercion (in other words, outside the rule of law), freedom would be preserved in an anarchist fashion (Anth, AA 07, p. 331).

Could Kant be saying, instead, that, exclusively by means of an instrumental use of reason, self-interest would lead devils to spontaneously unite under coercive laws to assure mutual safety? As already suggested, that is also highly doubtful. To begin with, this would lead us to conclude that Kant himself somehow took the source of juridical normativity to stem from the idea of prudence in the service of self-interest (which is not an accurate manner to reconstruct Kant's thought on the normativity of right). Beyond that, for evil beings who do not care about their mutual preservation under universal laws, self-interest alone, taken to its limits, would more probably bring forth barbarism, a highly unstable condition in which crude power would be lawless exerted at the expense of the freedom of many (Anth, AA 07, p. 331). In fact, according to Kant, even those "well-disposed and law-abiding" (MS, AA 06, p. 312) – which certainly is not the case of

[5] Admittedly, this is a different and maybe more radical depiction than the one Otfried Höffe presents in *'Even a Nation of Devils Need the State': the Dilemma of Natural Justice*. In his discussion, Höffe assumes a rational devil as a being who, "in his coexistence with others like him, allows himself to be guided by prudence alone" (HÖFFE, 1992, p. 125). Cooperation would be achieved only to assure "mutual advantage" (HÖFFE, 1992, p. 125). Outside that, devils would be "unhesitatingly tending towards dishonesty and deception" (HÖFFE, 1992, p. 125). I follow Höffe's account only partially. On one hand, it undoubtedly seems that, on a Kantian view, far from a someone who does the evil for the sake of evil, a devil is an archetype of an egotistical being who acts on the basis of a prudential reasoning alone (namely, in order to satisfy inclination). On the other hand, however, this supposed ability to cooperate for the sake of mutual advantage is not the defining trait that is being stressed out as a way of understanding what would be typical of a devil who *has evil intentions*. Kant's point can be read as emphasizing that devils not necessarily would cooperate precisely because they are willing to make an exception of themselves even when their preservation, as a group, is put under the scope of universal laws. They are rational, but evil. Now, if they are rational, then the *evilness* of their inner intentions is to be taken as entailing something to which they would rationally commit (RIPSTEIN, 2009, p. 108). Otherwise, it would be a matter of mere wishful thinking not to be taken that seriously (and devils would more likely be fools in Kant's discussion). Once this commitment is taken as a matter of actual prudential reasoning, the evilness of these devils, as well as the instrumental role of State in redirecting these devils' actions by coercive means (through inclinations), are the points to be discussed in *Toward Perpetual Peace*.

hypothetical devils – would still need a State equipped with precise and public adjudicatory rules set specifically to solve conflicts independently of unilateral conceptions of right.

As O'neill already pointed out, Kant's view on politics is, in many aspects, a "robustly realist political position" (O'NEILL, 2018, p. 220). His discussion about devils can clearly be seen as another instance of this political realism. One can understand this by paying attention to Kant's actual (but not exactly immediate) phrasing. Although usually assumed, without much questioning, that the *creation* or *stablishing* of State [*Staaterrichtung*] (ZeF, AA 08, p. 366) is, in itself, the (solvable) problem for a race of devils, there is another way to read the passage. It is also possible to understand that the real problem to be solved amounts exclusively to what Kant textually presents in quotes, right after initially saying that such problem (not yet specified by him) would be solvable even for a race of devils. Here is what Kant actually raises as a problem for a nation of devils:

> "To form a group of rational beings, which, as a group, require universal laws for their preservation, of which each member is, however, secretly inclined to make an exception of himself [*insgeheim sich davon auszunehmen geneigt ist*], and to organize them and arrange a constitution for them in such a way that, although they strive against each other in their private intentions [*Privatgesinnungen*], the latter check each other in such a way that the result in their public conduct is just as if they had no such evil intentions [*bösen Gesinnungen*]" (ZeF, AA 08, p. 366).

Focusing only on Kant's quotation above, the problem lies specifically in organizing rational and egoistical beings who, despite secretly willing to harm each other for the sake of the maximization of their own self-interest (precisely by making themselves an exception to a universal rule set for their preservation), should behave externally as if they were not inherently egoistical. In this sense, the very stablishing of the State *per se* is not the problem that Kant intends to address when these devils are brought to discussion, but, actually, part of the solution to the deeper problem posed by any attempt to regulate the conflictual relations of such evil beings. According to this reading, the task involved in coordinating evil beings under universal laws is, in fact, the actual problem that demands a solution, namely, the stablishing of State. Anarchy is not a plausible solution, because, from Kant's realistic point view, each devil would use the power of choice only to maximally satisfy his own interests. Again, there would not be "law and freedom without force" (Anth, AA 07, p. 331) but, instead, "force without freedom and law" (Anth, AA 07, p. 331), which amounts to barbarism (a condition which contradicts reason itself).

Despite mentioning the creation of State in the beginning of the passage, Kant *describes* and *formulates* the actual problem (without any reference to the creation of State itself) as being *only* a matter of coordinating evil beings who interact with each other and, presumably, would allow their inclinations alone to rule their uses of the power of choice seeking only to satisfy self-interest. Since this rational beings are evil, their reciprocal interactions would be maintained in absolute disregard for each other's freedom. According to Kant, however, a solution to said problem (and also the guarantee of perpetual peace from the rational point of view of humans beings) is possible, for it does not lie in "moral improvement" or "inner morality" (ZeF, AA 08, p. 367), but precisely in the proper use of sensible inclinations (the mechanical course of nature) to promote the rule of law which is juridically prescribed by reason [*rechtlichen Vorschrift*] (ZeF, AA 08, p. 367).

And how can this problem be solved in the case of a race of devils? Precisely by means of coercive positive laws. In short, Kant's point is that radically egoistical devils would certainly face major problems when forced to interact with each other (as humans historically do), but these problems,

even for them, as evil as their intentions might be, could be solved under a constitution and public positive laws. According to this interpretation of *Toward Perpetual Peace*, the actual stablishing of the State is not exactly the problem to be solved by devils - or, as William Clohesy prefers it, *for* devils (CLOHESY, 1995, p. 738). To put it in simpler terms, Kant's concern is not to discuss whether devils would be able or not to perform, out of their selfish reasons, the moral duty to leave the state of nature – *exeundum esse e statu naturali* (MS, AA 06, p. 312; RGV, AA 06, p. 97).

What Kant is actually proposing is that, despite invariably acting egoistically, devils would nevertheless submit, via coercion, to public and coercive laws designed to protect and promote freedom. As well known, reason itself commands each rational being to leave the state of nature as a matter of moral duty. Being inherently evil, devils would never leave the state of nature (they would prefer barbarism). However, although not out of the motive of duty itself, even devils would obey reason's command (from the point of view of a rational concept of right) if externally demanded by juridical laws to do so (regardless of their capacity or not to bring about this juridical condition by themselves). Preferring always to egoistically earn all benefits, they would without doubt *unwillingly* remain in a juridical condition in which each one's freedom is supposed to be preserved (and correspondently limited) according to universal laws. In other words, they would unite in this last manner – and this is the detail of the utmost importance not always emphasized enough – insofar as external coercion necessitates them, *per motiva*[6], to do so.

Juridical laws are supposed to operate on the level of the determining grounds of choice, as aversions (MS, AA 06, p. 219), to prevent such devils from externally acting against each other, as they would if led by their evil motivations alone. Inasmuch as avoiding the juridical consequences of violations would clearly meet the best interest of such beings, submission to juridical laws, even for a race of devils, also becomes a matter of self-interest (and, consequently, of inner motivation). And this is where the mechanical course of nature can be properly put to its use. Hence, Kant's not exactly surprising insistence on the fact that "what is of paramount importance in organizing the State well [...] is that the State directs the forces within it against each other in such a way that the one hinders or nullifies the destructive effects of the other" (ZeF, AA 08, p. 366).

Recognizing the pivotal role of public coercive laws in Kant's argument is crucial to a deeper understanding of the sense in which the mechanical course of nature can instrumentally serve

[6] In his *Lectures on Ethics*, discussing the concept of moral coercion, Kant mentions an important distinction between causes and motives as means by which someone can be necessitated to act. He starts by saying precisely that coercion in general is related to a "necessitation to action" ["*Nötigung zur Handlung*"] (V-Mo/Collins, AA 27, p. 266). The main feature of a *necessitation* is that it operates as a necessary condition for an action that wouldn't take place otherwise (V-Mo/Collins, AA 27, p. 266). But, just as we have two kinds of *arbitrium* (*brutum* and *liberum*), we also have, correspondently, two modes of coercion. One is the pathological coercion, by which one action is made necessary "*per stimulus*" (V-Mo/Collins, AA 27, p. 266). The other, a properly practical coercion, involves making necessary an action, unwillingly, but only "*per motiva*" (V-Mo/Collins, AA 27, p. 266). Due to *arbitrium liberum*, only this second kind of coercion is a necessitation in the case of human beings (non-rational animals, on the other hand, are pathologically coerced). Kant makes use of two examples to illustrate his point and to highlight the importance of the distinction between stimuli and motives as conditions for action. First, he considers the case of a stingy person who, although always preferring to earn as much benefits as possible from all situations, if faced with the unavoidable need to choose between two deals, would pick the most advantageous one motivated by his inclinations, even if this entails not getting the most benefits possible (which would occur only by closing both deals). Secondly, addressing the problem of torture, Kant says that even a person submitted to the cruelest acts would be capable of doing the contrary of what his torturer demands. In the case of humans, sensible stimuli alone do not necessitate in a pathological way, since the victim would not give in, unless led by a *motive of inclination* (to avoid, for example, the pain caused by the torturer). But the victim could endure the pain, also led by his inclinations, to protect a beloved relative for instance. Both stimuli are not enough unless taken as a motive to perform an action that would not occur without necessitation. According to Kant, even in the hard case of torture, for a human, one can "refrain from acting, independently of all sensible impulse" (V-Mo/Collins, AA 27, p. 267).

reason's ends towards a juridical condition. Especially from a teleological point of view, the interpretation via 'motivating mechanism' is correct, but only insofar as coercion is properly taken into account. And, as soon as coercion's crucial role in the arrangement of a community of evil beings under universal laws is correctly understood, an interesting feature of Kant's approach to the normativity of right is unveiled through the hypothesis formulated in *Towards Perpetual Peace*.

Kant asserts no more and no less than precisely this: problems of social coordination (again, under universal laws), even if they are unavoidable, *are still solvable* for a race of devils once the State manages to *redirect the forces within it* (by coercive means). The task may be difficult as it sounds, but, according to Kant's hypothesis, it can be accomplished within a well-organized State – and States in reality, although not perfectly organized, already show that (ZeF, AA 08, p. 366). He is not merely speculating whether it would be possible or not to organize devils in that manner under a republic. Kant is formulating a quite strong thesis: as evil as their inner motivations might be, devils would be coerced to be righteous and fair in their external and intersubjective relations (in fact, apparently only under coercive laws would it be possible to prevent them from relapsing into barbarism).

These devils would certainly not be "morally good" [ein moralisch-guter Mensch] (ZeF, AA 08, p. 366), but they would be, in Kant's own words, "good citizens" [guter Burger] (ZeF, AA 08, p. 366). Now, what is interesting about Kant's argument is that a *good citizen* is not taken, here, merely as someone who blindly obeys positive laws of any sort (as it is in its negative and more recent connotation). That becomes clear once we remember that Kant's focus, on the *First Supplement*, is the guarantee of the accomplishment of reason's ends with regard to a rule of law in which freedom is to be preserved under universal laws. Kant does not seem to raise the hypothesis of a nation of devils under any presupposition regarding the form of government that they would possibly accept. He does not state, for instance, that devils would only be capable of organizing themselves under a despotic rule of "law and force without freedom" (Anth, AA 07, p. 331). Kant is discussing the instrumental use of a mechanical course of nature specifically as a guarantee of a State under a republican Constitution – which is the only one "in perfect accordance with the right of humankind" (ZeF, AA 08, p. 366).

A good citizen, in Kant's scheme, is someone who acts in accordance with positive laws designed specifically to bring about reason's juridical prescription. In this sense, Kant's thesis is as plain and strong as it sounds: coercion, exerted properly and in an organized manner, under a republican Constitution, would necessitate even evil devils to be *good citizens* and to refrain from violating each other's freedom (in accordance with universal laws). So understood, the difference between a morally good person and a good citizen is put, at first glance, exclusively on the motivational level - and coercion, as a necessitation by motives for rational beings, is what distinguishably motivates devils to be good citizens (morally virtuous beings would respect each other's freedom out of the motive of duty itself).

However, if we are to understand how the hypothesis of righteous devils can be reconciled with Kant's own views on the normativity of right, then we need to interpret the passage in *Toward Perpetual Peace* in the light of two methodological precautions that the text itself seems to allow. First, it is not in question whether devils would be capable of spontaneously organizing under such State or not (since, once again, the stablishing of the State itself – e*xeundum esse e statu naturali* – is not Kant's actual problem). What must be presupposed, as Kant himself does, is that they would

be coerced to behave in certain ways by positive laws. Therefore, these citizen-devils would follow juridical laws with some regularity, despite doing that for the sake of their own selfish interests. Secondly, we are allowed to presuppose also within the textual limits set by Kant's hypothesis that, alongside its coercive laws, this Constitution would minimally satisfy Kant's republicanism. These devils would certainly not live under a perfect republic – which would be possible only for angels (ZeF, AA 08, p. 366) –, but neither would they live under a despotic State or, even worst, in a barbaric condition. In other words, what positive law prescribes, in this hypothetical scenario of a nation of devils, is also *right* according to *Kant's Doctrine of Right* (MS, AA 06, p. 229).

Given the way Kant presents the problem of a race of devils and its solution under the presupposition of coercive laws - in the teleological scope of a mechanical course of nature (ZeF, AA 08, p. 368), an interesting question arises: if even radically selfish beings could maintain righteous external relations out of a radical self-interest alone, what normative thesis about the juridical limitations imposed upon our maxims and actions can be drawn from the seemingly radical case of a selfish but still righteous devil? In other words, how can the extreme case of a righteous devil help us to understand what ought to be juridically demanded from rational beings under universal laws of freedom on a Kantian account of right? I'll deal with these questions in the next section.

3. The race of devils and the normativity of right

As already said, at first glance, Kant's distinction between a morally good person and a good citizen (which a devil could become) apparently would be simply bringing to light the underlying disjunction between what can be ethically demanded from a rational being in one hand, and, on the other hand, what can be juridically demanded from the same rational being. Basically, as subject to the external lawgiving of juridical laws, a rational being is not required to perform duties out of the motive of duty itself (contrary to what occurs, as well known, from an ethical perspective). This is precisely what the standard 'motivating mechanism' line of interpretation argues.

However, apart from this more obvious conclusion, isn't the criterion of right somehow connected to this motivational independence? In other words, isn't the very definition of a good citizen (i.e., someone who obeys the laws of a republican State), in *Toward Perpetual Peace,* dependent on a criterion of right that is, in itself, understood as such precisely because of its motivational independence?

Were Kant's remarks concerning this motivational independence of right confined to the discussion of devils in *Toward Perpetual Peace*, one would be tempted to immediately dismiss the question in favor of the 'motivating mechanism' reading. It so occurs, however, that the same motivational independence of right is highly emphasized by Kant in the *Introduction to the Doctrine of Right, §C*. In fact, it is quite remarkable that two out of the four paragraphs of this last section are dedicated to Kant concluding that the universal principle of right is the source of juridical obligation, but does not demand, at the same time, right actions also to be performed exclusively on the basis of this obligation (MS, AA 06, p. 231).

In this discussion, when the normativity of right itself is in focus, Kant explicitly says that, if a doctrine of right does not intend "to teach virtue, but only to set forth what is right, one may not and should not represent that law of right as itself the incentive to action" (MS, AA 06, p. 229). One would immediately and correctly argue that *not representing (or adopting) the law of right as an incentive to perform a right action* does not imply that the law of right itself (as the source of

normativity) is not built upon a motivational criterion for the evaluation of actions. Therefore, the objection would proceed, no normative conclusions are to be drawn from *§C* alone.

To answer this objection, it is important to remember that, if the universal principle of right is not demanded as the incentive for right actions, this occurs precisely because the principle itself, in the first place, is formulated in a way that makes the utter motives of the agent irrelevant. In fact, to solve the philosophical problem enunciated in *§B* (i.e. to stablish the criterion by which right can be recognized as such), Kant introduces a principle to which the incentive [*Triebfeder*] (MS, AA 06, p. 218) that drives the agent is irrelevant to the very criterion of right itself.

As well known, the universal principle of right is formulated exclusively in terms of the possibility of coexistence between an action and the freedom of choice of others (MS, AA 06, p. 230). In Kant's view, when it comes to a juridical evaluation of actions, it shall not be questioned whether the agent acts or not out of respect for the freedom of choice of others, but only if, by using his own free choice himself, he imposes any undue restriction that deems the use of free choice by others impossible under universal laws. In this manner, even the hypothetical devil of *Towards Perpetual Peace*, although not morally good, can be a good citizen – i.e., someone whose actions, according to a rational principle, are deemed right precisely by not impairing the freedom of others, even if only out of self-interest (or, more specifically, because State's coercion is exerted as means to necessitate even devils to refrain, out of self-interest, from violating each other's freedom).

To understand why the universal principle of right can be reconstructed in such manner, it is useful to remember that Kant operates under the most basic distinction between principles of execution [*principium executionis*] and principles of adjudication [*principium diiudicationis*] (ALMEIDA, 2006, p. 210; HÖFFE, 1989, p. 152; KLEIN, 2009, p. 67; PERES, 1998, p. 49). Georg Mohr and Allison remember that Kant draws the difference between those two kinds of principles in his *Lectures on Ethics* (MOHR, 2019, p. 75; ALLISON, 1990, p. 233). There, this difference is outlined on the basis of two questions that can be asked in regard to one's actions. One can ask, first, whether his actions are morally good or not (V-Mo/Collins, AA 27, p. 274), and then the problem is concerned exclusively with the evaluation of a given action as right or wrong. But one can ask, additionally, what motivates him to perform an action that is morally good. In this second case, the principle becomes a principle of execution (V-Mo/Collins, AA 27, p. 274).

This distinction is an important one because the criterion for right or wrong, by evaluating our actions in the scope of their compatibility with the freedom of choice of others, does not ask for the motivation of the agent who acts in that way. The universal principle of right is presented by Kant in the following manner: "Any action is right if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law" (MS, AA 06, p. 230).

It is easy to note that the possibility of coexistence between one's actions or maxims and the freedom of choice of others is not dependent on any particular motive adopted by an agent. Even Kant's reference to the agent's maxim in this context is not to be understood as implying a motivational reading of the universal principle of right. Indeed, nothing prevents a given maxim to be built upon incentives of sensuous nature (and not upon the motive of duty itself) and yet to be harmless in regard to others' freedom. The case of a devil whose maxims are done out of

fear of coercion are the best example. In Kant's radical hypothesis, devils would not harm each other's freedom exclusively at the prospect of avoiding the juridical consequences of violations[7].

Despite their wicked motives, they would still have to be deemed as fully righteous under a Kantian moral account of right. As Kant himself states, "anyone can be free so long as I do not impair his freedom by my external action, even though I am quite indifferent to his freedom or would like in my heart to infringe upon it" (MS, AA 06, p. 230)[8]. Devils, understood as radically egoistical and evil beings, seem to fit precisely this description.

Once the difference between principles of execution and principles of adjudication is correctly understood, it is also possible to conclude that the motivational independence of right is not restricted to the mere difference between the accomplishment of a duty out of the motive of duty itself, typical of ethical lawgiving (which demands the conformity with law in the form of morality), and the accomplishment of a duty on the basis of any other motives (empirical ones included), typical of juridical lawgiving (which demands conformity with law in the form of legality). The difference pertains to the very activity of reason. In its legislative task, from a juridical point of view, reason legislates specifically through a principle of adjudication alone (and this is the reason due to which legality becomes the typical form of conformity to laws admitted in a juridical domain).

It is possible to notice, therefore, two important aspects deeply related to the interpretation of the universal principle of right: i) the criterion of right is not built on an evaluation of agent's motivation, but takes in consideration exclusively the external dimension of his actions and maxims; ii) due to this circumstance, on the obligational level that underscores the normativity of right, the principle itself cannot be constructed as a principle of execution, under the penalty of contradicting the criterion of right itself (which does not ask for the agent's motivation).

This is decisive for our comprehension of a Kantian account of the normativity of right. Returning to the hypothesis of a race of devils, even if these rational beings do not determine their choice unless under motives of self-interest, their subjective principles of execution of actions would still survive the universalization test imposed by the universal principle of right (as long as no

[7] It is interesting to remember that, in the essay *'Devil's Apology'*, wrote in 1795, Johann Benjamin Erhard famously states that "full compliance with the laws of positive right is therefore no proof of a moral disposition because it can result from the fear of giving others an example of deviation" (ERHARD, 2019, p. 213). The deeper question raised by Kant's hypothesis of a nation of devils is not whether these evil creatures, as good citizens, would have a good moral disposition. It is granted that they do not, since they are evil. The actual problem is to understand whether they could be deemed righteous in their interpersonal relations, despite their complete lack of a good moral disposition. If they can, what does that tell us about Kant's views regarding the normativity of right?

[8] In the essay *On the Common Saying: This May be True in Theory, but It Does Not Hold in Practice*, similarly, Kant defines *right* as "the limitation of the freedom of others to the condition that it is consistent with mine in accordance with a general law" (TP, AA 08, p. 292). On the other hand, "public right (in a commonwealth) is merely the condition of a real [wirklichen] legislation in accordance with this principle and coupled with power" (TP, AA 08, p. 292). What is interesting is that this definition is presented by Kant in the context of his discussion about the principle of equality (alongside the principles of freedom and independence) that should ground a "civil condition" (TP, AA 08, p. 292). Now, a civil condition (which presupposes a *real legislation aligned to the principle of right as it is initially defined*) is nothing more than a "condition of equality of action and reaction of a mutually limiting choice in accordance with the general law of freedom" (TP, AA 08, p. 292). As one can clearly see, the concept of right (under the principle of equality) is conceived as a matter of a limitation of choice (power of maxims) in accordance with a universal law of freedom, or, more specifically, a law of external freedom. This limitation is demanded, however, only insofar as anyone's uses of choice, from a juridical point of view – at least a rational one grounded on the concept of right and on the principle of equality correspondently – cannot impose an undue restriction of the freedom of others, regardless of which may be each one's inner motives to not infringe upon other's freedom.

harm is brought to others). In other words, the universal principle of right admits a wider range of maxims, including those that are built exclusively on the basis of incentives of sensuous nature, imposing upon them only the restrictive condition according to which the freedom of others shall be left unimpaired according to universal laws.

This is possible, once again, because the criterion of right presented in the universal principle of right is deflated in motivational terms. And this motivational independence of right turns out to be the key to understanding the normativity of right as circumscribed specifically to the domain where actions and maxims of a rational agent can somehow influence other's freedom of choice. If devils, as evil as they might be, would still be righteous (provided, of course, that they are coerced by positive laws), that is due to the fact that the normativity of right, in its very rational source in practical reason, regulates our maxims and actions in a very particular manner.

The concept of right (the introduction of which precedes the formulation of the universal principle of right) can clarify this last point further. In *§B*, Kant exposes the three elements of this concept from the point of view of the juridical obligation related to a moral "concept of right" (MS, AA 06, p. 230): i) first, right deals exclusively with external actions of rational beings that can influence each other; ii) secondly, what matters, in this relation, is the power of choice of each rational being involved; iii) thirdly, right does not deal with the matter of choice, but with its form, i.e., with choice as a power that should be regarded as free.

Focusing on the first element, what is essential is that, unless an external behavior is brought about in a relation in which rational beings can influence each other, the juridical evaluation of an action as wrong or right is not set in place. Mental states of a given individual, for instance, are not relevant under a moral concept of right. But acts performed outside the domain of reciprocal relations are also irrelevant under this concept. Robinson Crusoe's acts were not juridically relevant (at least those perpetrated before he finally met Friday). Likewise, a suicidal hermit who, *ex hypothesis*, decides to seclude himself to put an end to his life would not have his acts evaluated from the rational point of view of a moral concept of right.

Beyond that, in the light of the second and third elements combined, the concept of right is related to the form of choice regarded merely as free. In the *Doctrine of Virtue*, Kant notes that "an end is a subject of free choice, the representation of which determines it to an action (by which the object is brought about)" (MS, AA 06, p. 385). In the same context, end-setting is outlined as "an act of the freedom on the part of the acting subject, not an effect of nature" (MS, AA 06, p. 385). And finally, reasserting what is typical of human choice, he says that the "capacity to set oneself an end – any end whatsoever – is what characterizes humanity (as distinguished from animality)" (MS, AA 06, p. 392). As Arthur Ripstein points out, humans are different from animals because they can "choose which ends to pursue" (RIPSTEIN, 2009, p. 362).

If Kant defines a free choice as the "capacity to set oneself an end – any end whatsoever" (MS, AA 06, p. 392), then this capacity is the form of choice. The matter of choice encompasses, on the other hand, more particular ends freely adopted by rational beings. The moral concept of right is not immediately concerned with these particular ends. Therefore, the evaluation of an action is concerned with the preservation of choice as a capacity that each rational being shall have to set and pursue his own ends.

In short, the concept of right imposes a formal limitation upon actions and maxims within the domain of external relations. This means that the moral concept of right does not say which ends shall be pursued by each one, but demands each end set by oneself to be limited to the conditions under which the freedom of choice can be used by everyone else according to universal laws. Since the matter of choice is not relevant, one can also conclude that the concept of right – and the universal principle of right – is not exactly concerned with the promotion of ethical ends (which ought to be done out of the motive of duty itself).

Consequently, even under this moral concept, the source of the normativity of right cannot be regarded as merely instrumental to the promotion of virtue, as some interpreters suggest (RILEY, 1982, p. 131), stemming, instead, from the compatibilization of external uses of the power of choice - regardless of which one's particular motives - by agents whose reciprocal coexistence in community is unavoidable (MS AA, 06, p. 307). The power of choice as a free one, not virtue is the aim of right from a rational point of view.

This obviously does not preclude the collateral role, on a political level, that each State has in promoting the "good moral education of a people" (ZeF, AA 08, p. 366), as Kant himself admits. But it shows that a moral approach of right is developed by Kant on normative grounds that are set independently of ethical considerations of this sort. This explains why, strictly under the moral normativity of right, even the radically egotistical devil described by Kant in *Toward Perpetual Peace*, although not a virtuous being, can still be righteous, as long as his actions do not impair the freedom of choice of others.

4. Concluding remarks

The interpretation above tries to show that, correctly read in the light of the central role of coercion as a mean to promote reason's prescription from a teleological point of view, the hypothesis of a race of devils can be developed and fully integrated into a Kantian account of the normativity of right (i.e., under a moral concept of right and a rational principle of pure practical reason).

It's worth noting, very briefly, that there are two most immediate advantages of this reading. First, it preserves the moral source of the normativity of right strictly according to Kant's division of the metaphysics of morals and avoids the 'prudence in the service of self-interest' interpretation. Secondly, and most importantly, the criterion by which positive laws shall be deemed as right or wrong under moral concepts and rational principles (MS AA, 06, p. 230) seems to also imply an evaluation regarding the State's capacity to regulate conflictual relations independently of each one's virtues and, what is of the utmost importance, without interfering in each individual's pursuit of happiness.

Finally, by touching upon that particular subject, one may immediately remember that, among other texts, Kant's criticism towards positive laws that interfere in this last domain is made explicit in the famous essay *On the Common Saying: This May be True in Theory, but It Does Not Hold in Practice*. By discussing the principle of political freedom, Kant says that a "paternalistic government" [*väterliches Regierung*] (TP, AA 08, p. 290) is the worst kind of "despotism" [*Despotismus*] (TP, AA 08, p. 291). This government would be despotic precisely by suppressing the freedom of its citizens to pursue their very own particular ends as if they were incapable of independent self-determination. Therefore, as Kant himself states, under civil laws, each is allowed to "pursue

happiness in the way that he sees fit, as long as he does not infringe on the freedom of others to pursue a similar end, which can coexist with the freedom of everyone" (TP, AA 08, p. 290).

In short, this reading is useful at least in showing that even Kant's moral account of right, although not a prudential one, presupposes a kind of normativity that not only is not concerned with setting which ends each one shall pursue, but that also cannot be taken as merely instrumental to the promotion of virtue. The moral concept of right and its corresponding principles are binding to rational beings from a moral point of view not at the prospect of an ethical improvement, but, instead, insofar as freedom of choice is also a rational demand for beings to whom external relations of reciprocal influence are unavoidable.

Bibliographic References

ALLISON, H. E. 1990. *Kant's theory of freedom*. Cambridge: Cambridge University Press.

ALMEIDA, G. A. 2006. Sobre o princípio e a lei universal do direito em Kant. *Kriterion,* Belo Horizonte, n. 114, p. 209-222, Dec 2006.

BRANDT, R. 2012. Kant as rebel against the social order. In: SHELL, S. (Ed.) M.; VELKLET, R. (Ed.). *Kant's observations and remarks: a critical guide.* Cambridge: Cambridge University Press.

CLOHESY, W. 1995. A constitution for a race of devils. In: ROBINSON, H. (Ed.). *8th International Kant Congress*. v. 2. Milwaukee: Marquette University Press.

ERHARD, J. B. 2019. 'Devil's apology'. Translators: James Clarke and Conny Rhode. *British Journal for the History of Philosophy*, London, v. 27, n. 1, p. 194-215.

FLIKSCHUH, K. 2000. *Kant and modern political philosophy*. Cambridge: Cambridge University Press.

HÖFFE, O. 1992. 'Even a nation of devils needs the State': the dilemma of natural justice. In: WILLIAM, H. (Ed.). *Essays on Kant's political philosophy*. Cardiff: University of Walle Press.

HÖFFE, O. 1989. Kant's principle of justice as categorical imperative of law. In: YOVEL, Y. (Ed.). *Kant's practical philosophy reconsidered*: papers presented at the seventh Jerusalem philosophical encounter. Berlim: Springer Science + Business Media.

KANT, I. 2017. *A metafísica dos costumes*. 3. ed. Lisboa: Fundação Calouste Gulbenkian.

KANT, I. 1917. Anthropologie in pragmatischer Hinsicht. In: REIMER, G. (Ed.). *Kant's gesammelte Schriften*. v. 7. Berlin: Königlich Preussischen Akademie der Wissenschaften. (Anth, AA 07).

KANT, I. 2006. *Antropologia de um ponto de vista pragmático*. São Paulo: Iluminuras.

KANT, I. 2020. *À paz perpétua*: um projeto filosófico. Petrópolis: Vozes.

KANT, I. 1923. Das mag in der Theorie richtig sein, taugt aber nichts für die Praxis. In: REIMER, G. (Ed.). *Kant's gesammelte Schriften*. v. 8. Berlin: Königlich Preussischen Akademie der Wissenschaften. (TP, AA 08).

KANT, I. 1914. Die Metaphysik der Sitten. In: REIMER, G. (Ed.). *Kant's gesammelte Schriften*. v. 6. Berlin: Königlich Preussischen Akademie der Wissenschaften. (MS, AA 06).

KANT, I. 2009. *Fundamentação da metafísica dos costumes.* São Paulo: Barcarolla; Discurso Editorial.

KANT, I. 1911. Grundlegung zur Metaphysik der Sitten. In: REIMER, G. (Ed.). *Kant's gesammelte Schriften*. v. 4. Berlin: Königlich Preussischen Akademie der Wissenschaften. (GMS, AA 04).

KANT, I. 1974. Kants Vorlesungen. In: REIMER, G. (Ed.). *Kant's gesammelte Schriften*. v. 27. Berlin: Königlich Preussischen Akademie der Wissenschaften. (V-Mo/Collins, AA 27).

KANT, I. 2018. *Lições de ética*. São Paulo: Editora Unesp.

KANT, I. 2006. On the common saying: this may be true in theory, but it does not hold in prac-

tice. In: KLEINGELD, P. (Ed.). *Toward perpetual peace and other writings on politics, peace, and history*. London: Yale University Press.

KANT, I. 2014. *Princípios metafísicos da doutrina do direito*. São Paulo: Martins Fontes.

KANT, I. 1914. Religion innerhalb der Grenzen der bloßen Vernunft. In: REIMER, G. (Ed.). *Kant's gesammelte Schriften*. v. 6. Berlin: Königlich Preussischen Akademie der Wissenschaften. (RGV, AA 06).

KANT, I. 1998. *Religion within the boundaries of mere reason*: and other writings. Cambridge: Cambridge University Press.

KANT, I. 1996. *The metaphysics of morals*. Cambridge: Cambridge University Press.

KANT, I. 2006. Toward perpetual peace: a philosophical sketch. In: KLEINGELD, P. (Ed.). *Toward perpetual peace and other writings on politics, peace, and history*. London: Yale University Press.

KANT, I. 1923. Zum ewigen Frieden: ein philosophischer Entwurf. In: REIMER, G. (Ed.). *Kant's gesammelte Schriften*. v. 8. Berlin: Königlich Preussischen Akademie der Wissenschaften. (ZeF, AA 08).

KERSTING, W. 1992. Politics, freedom and order: Kant's political philosophy. In: GUYER, P. (Ed.). *The Cambridge companion to Kant*. Cambridge: Cambridge University Press.

KLEIN, J. T. 2009. O conceito kantiano de metafísica dos costumes. *Peri*, Florianópolis, v. 1, n. 1, p. 57-72.

KLEINGELD, P. 2004. Approaching perpetual peace: Kant's defence of a league of states and his ideal of a world federation. *European Journal of Philosophy*, v. 12, n. 3, p. 304-325.

KORSGAARD, C. M. 1996. *The sources of normativity*. Cambridge: Cambridge University Press.

LUDWIG, B. 2002. Whence public right?: the role of theoretical and practical reasoning in Kant's doctrine of right. In: TIMMONS, M. (Ed.). *Kant's metaphysics of morals*: interpretative essays. New York: Oxford University Press.

MOHR, G. 2019. Autonomy and moral empiricism: Kant's criticism of sentimentalist moral principles. In: BACIN, S. (Ed.); SENSEN, O. (Ed.). *The emergency of autonomy in Kant's moral philosophy*. Cambridge: Cambridge University Press.

O'NEILL, O. 2018. *From principles to practice*. Cambridge: Cambridge University Press.

PERES, D. T. 1998. Imperativo categórico e doutrina do direito. *Cadernos de Filosofia Alemã*, São Paulo, n. 4, p. 43-64.

RILEY, P. 1982. *Will and political legitimacy*: a critical exposition of social contract theory in Hobbes, Locke, Rousseau, Kant and Hegel. Cambridge: Harvard University Press.

RIPSTEIN, A. 2009. *Force and freedom*: Kant's legal and political philosophy. Cambridge: Harvard University Press.

WOOD, A. 1999. *Kant's ethical thought*. Cambridge: Cambridge University Press.

WOOD, A. 2002. The final form of Kant's practical philosophy. In: TIMMONS, M. (Ed.). *Kant's metaphysics of morals*: interpretative essays. New York: Oxford University Press.

Kant and the prominent tone of superiority

Kant e o enaltecido tom de superioridade

Frank Rettweiler
Universität Vechta
frank.rettweiler@posteo.de

Abstract: The paper discusses Kant's views on the alleged tone of superiority in philosophy, a topic which the philosopher directly addressed in the paper called *On a recently prominent tone of superiority in philosophy*. It attempts to provide a somewhat systematic approach regarding this particular subject, while also connecting it to pivotal themes in Kant's philosophy, like philosophy as labor, enlightenment, knowledge (and wisdom) within the limits of theoretical reason and perpetual peace.

Keywords: Enlightenment; Kant; philosophical labor; knowledge; superiority; wisdom.

Resumo: Este artigo discute a percepção de Kant sobre o suposto tom de superioridade na filosofia, um tema que o filósofo abordou diretamente no artigo intitulado *Sobre um recentemente enaltecido tom de distinção na filosofia*. Pretende-se apresentar uma abordagem sistemática do tópico, conectando-o a temas centrais da filosofia kantiana, como filosofia como trabalho, esclarecimento, conhecimento (e sabedoria) nos limites da razão teórica e da paz perpétua.

Palavras-chave: Iluminismo; Kant; trabalho filosófico; conhecimento; superioridade; sabedoria.

Recebido em 10 de setembro de 2025. Aceito em 12 de novembro de 2025.

"*Vornehm*"

Jacob and Wilhelm Grimm point out in their dictionary that the adjective "*vornehm*" was narrowed in their time to the meaning of an advantage through birth and rank (VORNEHM, 2025). We are also dealing with this narrowing of the meaning of "*vornehm*" in the texts of Kant, which I want to discuss. Young people today whose native language is German still understand this adjective in this meaning, but it is rarely actively used anymore. In the 1960s and 1970s, this word was still used more often in the sense of "*Vornehm-Tun*", which means something like pretending to be superior or noble. Someone "*tut vornehm*" (acts noble) if this person at least temporarily displays the behavior of an upper social class (a more noble class), even though he or she does not belong to this class. The behavior of "*vornehm tun*" also includes the attitude of not wanting to do menial work or of thinking oneself superior to it. In this sense, there is something arrogant and presumptuous about "*vornehm tun*".

In Kant's time, when society was much more hierarchically structured than it is in Germany today, there was nothing offensive about saying that a person was "*vornehm*" (noble). She was then usually a noble or aristocratic person. In this respect, one can say of a person in Kant's time that they were "*vornehm*" (noble or superior) without wanting to say anything negative about them. The use of the adjective "*vornehm*" becomes problematic in philosophy, which is about the giving and taking of reasons. The reasons are important and not the person who presents them: it is about reasons regardless of the person. A person's status and rank are irrelevant, and their social class does not matter. That's why you can't be "*vornehm*" (noble or aristocratic) in philosophy. But you can pretend to be superior or noble ("*Vornehm tun*") in philosophy, precisely when you believe that you can rise above the give and take of reasons because you believe you have privileged access to the truth. The "*vornehme Ton*" (tone of superiority) has no place in philosophy.

And this is precisely what Kant wants to demonstrate in his essay *On a recently prominent tone of superiority in philosophy.* He argues not only against Schlosser, but against an entire group of philosophers who, while claiming to be enlightened, used their superior tone to absolve themselves of philosophical work and thus undermined the Enlightenment. Here, Kant's line of thought will be traced, which is clearly directed against the refined tone in philosophy, because the latter believes it can dispense with elaborate argumentation in favor of intuitive insight.

The addressees of Kant's essays of 1796

In his essay *Von einem neuerdings erhobenen vornehmen Ton in der Philosophie* (*On a recently prominent tone of superiority in philosophy*)[1], Kant does not specify who raised this tone of superiority. In his essay *Kants Kniefall vor der verschleierten Isis*, Norbert Klatt quotes a letter from Johann Georg Schlosser to Johann Georg Jacobi, in which Schlosser sees himself as the addressee of Kant's essay from the *Berlinische Monatschrift* (May of 1796). Schlosser, who married Goethe's sister Cornelia, was also in contact with Count Friedrich Leopold zu Stolberg. Johann Georg Jacobi, in turn, was the brother of the philosopher Friedrich Heinrich Jacobi. Princess Amalie von Gallitzin, in whose garden the Königsberg philosopher Johann Georg Hamann was buried, also belonged to this circle. She maintained contact with the Dutch philosopher Frans Hemsterhuis, who in turn plays an important role in Friedrich Heinrich Jacobi's book on Spinoza. This circle can

[1] I use the English translation of *The Cambridge Edition of the Works of Immanuel Kant*, edited by Henry Allison and Peter Heath (2002).

be assigned to the literary movement of *Empfindsamkeit* (sentimentalism), in which an intimate relationship with God should be cultivated[2].

Even if Kant does not name a specific addressee in his essay on the "tone of superiority", this circle and Schlosser in particular can be identified as the addressee. In his essay from December of 1796 (also published in the *Berlinische Monatschrift*), called *Verkündigung des nahen Abschlusses eines Tractats zum ewigen Frieden in der Philosophie* (*Proclamation of the imminent conclusion of a treaty of perpetual peace in philosophy*), Kant also explicitly refers to Schlosser (Kant AA 8:419). With this essay, Kant responded to Schlosser's reply to his essay on the "tone of superiority".

Philosophy as labor

"Philosophy" as a title for an activity is also used "as a decorative title for the understanding possessed by uncommon thinkers" (Kant AA 8:389/CE 431). And recently thinkers have emerged who are described by Kant as "*philosophus per inspirationem*" (Kant AA 8:389/CE 431), who cannot communicate the wisdom or knowledge they have acquired. The theoretical use of reason is denied knowledge of the supersensible. Such insight would be based on either a discursive or intuitive mind/understanding. The intuitive mind/understanding, which would "grasp and present the object immediately, and all at once" (Kant AA 8:389/CE 431) by means of an intellectual intuition, is superior to the discursive mind, because the latter has to work through concepts to develop its knowledge little by little, for what purpose sensual perception must first provide him with material. According to Kant, it can be explained "by the naturally self-seeking tendency in man" (Kant AA 8:389/CE 431) that an inclination arises to attribute to oneself an intuitive understanding in order to look down contemptuously on those who only have a discursive one, which they can use to gain knowledge. Anyone who attributes to themselves an intuitive understanding with which they believe they can acquire knowledge of the supernatural is already adopting an attitude that can be described as "*vornehm*" (noble or superior). And since we know that we do not have such an intuitive understanding, such a person is one who "*vornehm tut*" (pretends to be superior).

Kant distinguishes those "who *have enough to live on*" (Kant AA 8:390/CE 431) from those who have to work. And those who don't have to work think of themselves as "Vornehme" (nobles or superiors or aristocrats). Knowledge is something you have to work for, but the type of "philosopher of *intuition*" (Kant AA 8:390/CE 432) who think that they only have to listen to an "oracle within" (themselves) (Kant AA 8:390/CE 431) in order to participate in the whole knowledge that philosophers strive for.

> things have lately gone so far that an alleged philosophy is openly proclaimed to the public, in which one does not have to work, but need only hearken and attend to the oracle within, in order to gain complete possession of all the wisdom to which philosophy aspires (Kant AA 8:390 / CE 431).

The origins of the tone of superiority in the ancient philosophy of mathematics

In Kant's interpretation of Plato's philosophy, he asked himself how we can have a priori knowledge that goes beyond our *a priori* concepts? He saw that we have such knowledge in mathematics. After Kant, Plato already asked himself the question: "How are synthetic propositions possible

[2] I don't want to provide a definition of the literary movement of *Empfindsamkeit*, but rather just say that this circle around the Princess von Gallitzin, the Count of Stolberg and the Jacobi brothers is just a branch of the German-speaking *Empfindsamkeit.*

a priori?" (Kant AA 8:391 Footnote/CE 433). Since Plato did not consider that there was such a thing as pure sensory intuition, he assumed that our mind would have intuitions (ideas), but in our life, when the soul is bound in a body, we can only access them via "an intuiting of copies (*ectypa*)" (Kant AA 8:391/CE 432). For Kant, freeing oneself from physical restrictions in order to grasp the archetypes (ideas) themselves is the beginning of the enthusiasm for which Plato "put the torch" (Kant AA 8:392/CE 432).

Aristotle receives a different assessment from Kant: "The philosophy of Aristotle, on the other hand, is work" (Kant AA 8:393/CE 434). Aristotle made the mistake of extending the application of his categories to the supersensible, but the crucial difference is that for him philosophy has not to fall into a tone of superiority, but was work. It does not rely on a special cognitive faculty that we do not have, but is limited to giving and taking reasons.

The tone of superiority in the modern philosophy

Anyone who philosophizes – and this also applies to superior people – enters a field of "civil equality" (Kant AA 8:394/CE 415). Here Kant distinguishes "philosophizing and making philosophers" (Kant AA 8:394/CE 435): "The latter happens in the tone of superiority, if despotism over the reason of the people (and over one's own reason), by fettering it to a blind belief, is given out as philosophy" (Kant AA 8:394 Footnote/CE 435). Kant sees the tone of superiority within philosophy as a danger for philosophy itself.

No one should expect indulgence from those who raise a tone of superiority within philosophy, because in philosophy there is "freedom and equality in matters of mere reason" (Kant AA 8:394/CE 435). The person who refers to his feelings – or even "higher feelings" (CE 435) – In his philosophical argument also uses a tone of superiority, "for who will dispute my feelings with me" (Kant AA 8:395/CE 435). By invoking a feeling that is only accessible privately, one evades the public exchange of reasons and claims privileged access to particular reasons. Kant rejects a form of argument that claims to grasp an object through a feeling and pretends to be above communicating it through concepts. Taking up the legal diction of the *Critique of Pure Reason*, Kant says that the legality of the title of possession that one has acquired through feeling must first be proven. In the *Critique of Pure Reason*, no ability that we humans have that could prove the legitimacy of such knowledge could be identified.

However, the note in the next page (Kant AA 8:395/CE 436) makes it clear that in rejecting the appeal to feelings, Kant is concerned not only with the theoretical but also with the practical use of reason. Here it is important for him to distinguish between a pathological and a moral feeling as the determining factor of the will. The moral feeling does not precede the moral law, but follows from it, while the pathological feeling as the determining factor of the will can never be moral. We are then dealing with a material determining ground of the will and from this - Kant emphasizes this again clearly - only maxims of happiness can follow. In practical philosophy, appealing to a feeling does not, in principle, lead to a practical law.

The hunch/intuition (*Ahnung*)

Kant had already written in the note in Kant AA 8:394/CE 435 that the tone of superiority in philosophy will annihilate it through "by obscuration" (Kant AA 8:394/CE 435). Kant takes up this idea of the end or death of philosophy through the tone of superiority again when he talks

about the division of what is believed to be true into knowledge, belief and opinion, because his opponents of the tone of superiority want to expand this division to include the "*intuition* of the super-sensible" (Kant AA 8:397/CE438). In relying on such an *intuition* or hunch there is "an overleap (*salto mortale*) from concepts to the unthinkable" (Kant AA 8:398/438). The call to perform a *salto mortale* in philosophy goes back to Friedrich Heinrich Jacobi, who called for such a *salto mortale* from pure, rational, argumentative philosophy to faith in opposition to Spinoza in a conversation with Lessing.

Whether Kant was alluding directly to Friedrich Heinrich Jacobi or merely using the same term *salto mortale* cannot be said with absolute certainty, but the *intuition* or hunch as the basis for knowledge is in poor condition.

> For intimation is obscure expectation, and contains the hope of a solution, though in matters of reason this is possible only through concepts; if these are transcendent, therefore, and can lead to no true knowledge of the object, they must necessarily promise a surrogate thereof, supernatural information (mystical illumination): which is then the death of all philosophy. (Kant AA 8:398/CE 438).

The neoplatonism of Schlosser

Plato, as the author of the dialogues, is described by Kant as "Plato the academic" (Kant AA 8:398/CE 438) and although he was "the father of all enthusiasm *by the way of philosophy*" (Kant AA 8:398/CE 438), he is nevertheless acquitted of guilt. He distinguishes this academic Plato from the Plato who wrote the letters. Johann Georg Schlosser translated these letters from Greek into German. Kant cites quotations from the book with the new translation of the letters and comments on them:

> Who can fail to see here the mystagogue, who not only raves on his own behalf, but is simultaneously the founder of a club, and in speaking to his adepts, rather than to the people (meaming all the uninitiated), plays the *superior* with his alleged philosophy! (Kant AA 8:399/CE 439)

In the theoretical use of reason, this leads to a theophany, which leads to idolatry. Worship of God becomes superstition. In practical use, feelings are foisted on reason which can only paralyze practical reason.

But the philosophers of feeling believe that the philosophy of feeling would make us better. Now, according to Kant, one can see that an action was good, but how much of this action can really be attributed to a moral attitude cannot be said on an empirical basis. We can only say whether an action was morally good if we know from which maxim it was carried out and whether this maxim can be generalized. Kant speaks here about the possibility of universalizability because this requirement cannot be met by feeling as a basis for practical rules. The philosopher of feeling wants to derive a practical principle from feeling and see reason used to support this feeling. In doing so, he overturns the relationship between feeling and reason on its head. His daring attempt at a *salto mortale* failed. With emphatic words, Kant reminds us that we can hear the inner voice of reason and follow it, even if all our inclinations and considerations of advantage speak against it.

> What is it in me which brings it about that I can sacrifice the innermost allurements of my instincts, and all wishes that proceed from my nature, to a law which promises me no compensating advantage, and threatens no loss on its violation; a law, indeed, which I respect the more intimately, the more strictly it ordains, and the less it offers for doing so? (Kant AA 8:402/CE 442)

It is this question that, according to Kant, shows us "the magnitude and sublimity of the inward disposition in mankind" (Kant AA 8:402/CE 442) and creates in us a feeling that can make us morally better. It is a feeling that is not the foundation of reason, but is based on reason.

Philosophy as work/labor and the defense of the Enlightenment

Here, Kant claims, Archimedes found his point "on which reason can apply its lever" (Kant AA 8:403/CE 442). This Archimedean point is found through philosophical work. To do this, it must be examined what the concepts of the understanding are and how far they extend. And to do this, the faculty of reason must be examined, including its possible theoretical and practical uses.

Kant suspects that the use of the tone of superiority represents a general attack on the Enlightenment. And this is the reason why he raises his voice against the tone of superiority.

> To ensure such a claim did not strike me as superfluous at the present time, when adornment with the title of philosophy has become a matter of fashion, and the philosopher of vision (if we allow such a person) might – seeing how easy it is, by an audacious stroke, to attain without trouble to the summit of insight – be able unawares (since audacity is catching) to assemble a large following about him: […]" (Kant AA 8:403/CE 443).

The bold visionary and his followers look down with disdain on the school-like form of academic philosophical work that clings so closely to the formal. The accusation against them is that they are a "pattern-factory" (*Formgebungsmanufactur*) (Kant AA 8:404/CE 443) and that they are completely replacing them.

Kant reminds us that the concentration on form, in theoretical philosophy as the study of the forms of intuitions and forms of thought, was the first to explain how synthetic judgments are possible *a priori*. Reason leads us into the supersensible and can only ensure the practical reality of the objects of the supersensible through its practical use. But this insight is also based on practical laws that only consider the generalizable form of the maxims.

Kant once again emphasizes that there is something livelier in this formal observation than there is in the bold, visionary observation of the superior (*Vornehmen*).

> In both fields (theoretical and practical) it is not an arbitrary form-giving undertaken by design, or even machine-made (on behalf of the state), but above all a piece of handwork, dealing with the given object, and indeed with no thought of taking up and evaluating the preceding industrious and careful work of the subject, his own faculty (of reason); by contrast, the gentleman who opens up an oracle for the vision of the super-sensible will be unable to deny having contrived it by a mechanical manipulation of men's brain, and attached the name of philosophy to it for honorific purposes alone (Kant AA 8:404/CE 444).

Perpetual Peace in Philosophy

But Kant also offers a conciliatory tone in his essay *On a recently prominent tone of superiority in philosophy*, since both parties only want to make people righteous. The veiled goddess Isis, who is worshiped by the successors of the letter writer Plato, can be seen as the moral law within us and then Kant would also bow the knee to her. For didactic reasons, however, Kant prefers to explain the moral law "by logical instruction" (Kant AA 8:405/CE 444) and only then to personify it in order to give the moral law "an *aesthetic* way of presenting" (Kant AA 8:405/CE 444). However, he does not fail to point out the danger of an enthusiasm triggered by this.

In his essay *Proclamation of the imminent conclusion of a treaty of perpetual peace in philosophy published* in the *Berlinische Monatsschrift* in December of 1796, Kant once again shows the reasons why he considers such a peace in philosophy to be possible. In the preface to the first edition of the Critique of Pure Reason, published in 1781, Kant wrote: "The battlefield of these endless controversies is called metaphysics" (Kant KrV A viii/Kant AA 4:7/CE 99). But the *Critique of Pure Reason* has also laid the foundations for peace to come where previously endless controversies caused turmoil. But the fact that such a philosophical dispute has its positive sides is illustrated

by Kant's tendency "to *squabble* on behalf of one's philosophy" (Kant AA 8:414/CE 453). As he says "[...] this itch, I say, or rather *drive*, will have to be viewed as one of the beneficent and wise arrangements of Nature, whereby she seeks to protect man from the great misfortune of decaying in the living flesh" (Kant AA 8:414/CE 453).

At first glance this may sound ironic, but it is actually meant seriously. At the beginning of the essay, Kant had said, quoting Chrysippus: "Nature has given the pig a *soul*, instead of *salt*" (Kant AA 8:413/CE 453). This soul, later called life force, prevents the pig from rotting or decaying in the living flesh. In contrast to animals, we have gained self-consciousness and so – "in virtue of which" (Kant AA 8:414/CE 453) – have become an animal that has reason. However, through the power of reason we get into what Kant calls "trifling (*Vernünfteln*)" (Kant AA 8:414). We get into "trifling" because of the structure of reason, because it compels us to ask questions that we cannot answer theoretically. The preface to the first edition of the *Critique of Pure Reason* begins with exactly this:

> Human reason has the peculiar fate in one species of its cognitions that it is burdened with questions which cannot dismiss, since they are given to it as problems by the nature of reason itself, but which it also cannot answer, since they transcend every capacity of human reason(Kant KrV A vii/ Kant 4:7/CE 99).

We try to answer these irrefutable questions about the immortality of the soul, about the freedom and the existence of God, which arise from the nature of reason, within theoretical philosophy and thus entangle ourselves in contradictions so that metaphysics becomes that battleground, which Kant now thinks he can pacify.

Only through the practical use of reason can ideas acquire a moral and practical reality. To do this, however, it had to be shown beforehand that the theoretical use of reason could not give these ideas objective reality. It could only be shown in theoretical use that the ideas of freedom, immortality and God are conceivable without contradiction. The starting point for the moral and practical reality of these ideas is the idea of freedom in connection with the categorical imperative. Kant outlines this line of thought from the *Critique of Practical Reason* again in his essay from 1796:

> But now there actually is something in human reason, which can be known to us by no experience, and yet proves its reality and truth in effects that are presentable in experience, and thus can also (by an a priori principle, indeed) be absolutely commanded. This is the concept of freedom, and of the law that derives from this, of the categorical, i.e., absolutely commanding, imperative. Through this we acquire Ideas, that would be utterly empty for merely speculative reason, though the latter inevitably point us towards them as cognitive grounds of our ultimate purpose – and admittedly only moral and practical reality; namely, so to conduct ourselves as if we were given the objects of these Ideas (God and immortality), which may therefore be postulated in this (practical) respect (Kant AA 8:416/CE 455).

Through the systematic insight into the limits of theoretical knowledge, the practical use of reason opens up the basis for the assumption that we are free beings and, as a result, we can postulate the immortality of the soul and the existence of God. For Kant, the "impotence, on the one hand, of *theoretical proof*" (Kant AA 8:416/CE 455) and the "strength of the practical grounds" (Kant AA 8:416/CE 455) gives rise to the prospect of peace, which, however, does not let us rot because our psychological forces will be kept active through repeated attacks, which arise from the fact that the theoretical illusion that arises from the nature of our reason will remain in place.

Philosophy as a doctrine of knowledge and as wisdom

Kant does not want to accuse Schlosser of any bad intentions, as he believes that Schlosser also advocates "a mind attuned to promotion of the good" (Kant AA 8:419/CE 458). However,

Schlosser believes that he can get rid of philosophy as a doctrine of knowledge or skip it in order to move straight on to philosophy as a doctrine of wisdom. This also shows his superior way of thinking, from which the tone of superiority probably comes from. It was too much trouble and too much work for him to study critical philosophy, so "he has only looked at the final results proceeding from it" (Kant AA 8:419/CE 458). And Schlosser didn't like these results: "and so, without having first gone to school himself, he forthwith became the teacher of 'a young man who (he says) wanted to study the critical philosophy,' in order to advise him against doing so" (Kant AA 8:419/CE 458). This young man, if he follows his teacher and does not submit to the effort and work of school, is then easily a victim of "the art of persuasion, on subjective grounds of approval" (Kant AA 8:420/CE 458), instead of making judgments based on objective grounds. Anyone who follows the art of persuasion is satisfied with the "*semblance of truth*" (Kant AA 8:420/CE 458) and passes it off for "probability" (Kant AA 8:420/CE 458). But in the field of a priori knowledge there can be no probability at all.

As far as the scope of knowledge in the theoretical area is concerned, philosophy as a theory of knowledge only has a "limiting pretensions" (Kant AA 8:420/CE 458). Philosophy as a teaching of wisdom must always keep these limitations of knowledge in mind.

Anyone who imposes subjective reasons on the semblance of truth where objective reasons striving for truth can be had is violating what Kant calls the "duty of truthfulness" (Kant AA 8:421/CE 459). One can make mistakes if one wants to fulfill this duty. But anyone who is wrong is not therefore untruthful. Anyone who is guilty of dereliction of duty to truthfulness is a liar. For Kant, lying is the original sin, as he makes clear again at the end of his essay:

> The *lie* ("from the father of lies, whence all evil in the world hath come") is the truly vile spot in human nature, […] The commandment: *Thou shalt not lie* (were it even with the most pious intentions), if most sincerely adopted into philosophy, as a doctrine of wisdom, would alone be able, not only to procure eternal peace therein, but also assure it for all time to come. (Kant AA 8:422 / CE 459).

Bibliographic References

KANT, I. 2002-. *The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press.

KANT, I. 1900-. *Gesammelte Schriften*. Königlich Preussische Akademie der Wissenschaften (and successors) (Eds.). Berlin: de Gruyter (and predecessors).

KLATT, N. 1985. Kants Kniefall vor der verschleierten Isis. *Zeitschrift für Religions- du Geistesgeschichte*, [*s.l.*] v. 37, n. 2, p. 97–117.

VORNEHM. 2025. In: *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*. Trier: Trier Center for Digital Humanities. Available at: https://woerterbuchnetz.de/?sigle=DWB&lemid=V14132. Access: 09 sep. 2025.

Rationalizing and social irrationality from a Kantian perspective

Racionalização e irracionalidade social numa perspectiva kantiana

Eduardo de Oliveira da Costa[1]
Universidade Federal de Santa Catarina/Universität Vechta
eduardodeoliveiradacosta1999@gmail.com

Abstract: The present paper addresses the relation between, on the one hand, Kant's concept of rationalizing <*Vernünfteln*> and, on the other hand, the concept and phenomenon of social irrationality understood in a Kantian perspective. More specifically, this work argues that one can see in Kant's concept of rationalizing, which is understood by Kant as a form of irrationality, a source of social irrationality and, therefore, a conceptual tool that can help us to address the phenomenon of social irrationality in our society. To accomplish this task, this paper is divided in four sections: an introduction (1), a definition of the concepts of irrationality and social irrationality (2), an explanation of the essential characteristics of Kant's concept of rationalizing as one can find in Kant's works and in the interpreters (3), and a demonstration of the usefulness of Kant's concept of rationalizing for understanding the sources of social irrationality in our society (4).

Keywords: Kant; rationalizing; social irrationality; irrationality; rationality; ideology.

Resumo: O presente artigo aborda a relação entre, de um lado, o conceito de racionalização <*Vernünfteln*> de Kant, de outro lado, o conceito e fenômeno da irracionalidade social compreendido desde uma perspectiva kantiana. Mais especificamente, argumenta-se que se pode identificar no conceito de racionalização, o qual é entendido por Kant como uma forma de irracionalidade, uma fonte de irracionalidade social e, portanto, uma ferramenta conceitual que pode nos ajudar a abordar o fenômeno da irracionalidade social em nossa sociedade. Para cumprir com esse objetivo, este artigo está dividido em quatro seções: uma introdução (1), uma definição dos conceitos de irracionalidade e irracionalidade social (2), uma explicação das características essenciais do conceito de racionalização de Kant tal como elas são encontradas nas obras de Kant e nos(as) intérpretes (3) e uma demonstração da utilidade do conceito kantiano de racionalização para a identificação das fontes de irracionalidade social (4).

Palavras-chave: Kant; racionalização; irracionalidade social; irracionalidade; racionalidade; ideologia.

[1] This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Recebido em 31 de julho de 2025. Aceito em 14 de outubro de 2025.

1. Introduction

The concepts of social *rationality* and *irrationality*, when analyzed from the perspective of a Kantian social philosophy, have proven to be increasingly useful for dealing with the tension between rationality and irrationality that permeates social institutions. At the same time, Kant's concept of *rationalizing* <*Vernünfteln*> has been increasingly addressed by some of the interpreters of Kant's philosophy in the last decade[2]. Despite the distinctions on the way these different approaches understand the meaning of this concept and its importance within Kant's philosophical work, it is a common place to state that Kant understood the concept of rationalizing as denoting a kind of irrationality, i. e., for instance, "a use of reason that misses its final end, partly from inability, partly from an inappropriate viewpoint" (Anth, AA 07: 200) or even "only an empty use of reason which contains nothing in regard to the true ends." (V-Anth/Busolt, AA 25: 1481) In other words, the idea that a Kantian definition of irrationality encompasses Kant's concept of rationalizing does not seem hard to be defended.

On the other hand, the possibility of seeing Kant's conception of *rationalizing not only* as important for understanding irrationality in general and its sources from a Kantian perspective, *but also* as an important conceptual tool for understanding the sources of social irrationality is a much harder task. In other words, it is not exactly clear whether it is possible to explain social irrationality or, at least, one of its sources, using Kant's concept of rationalizing. The reason for this is that, although Kant's concept of rationalizing clearly fits into the concept of irrationality, the concept of social irrationality not only encompasses or deals with the concept of irrationality, but is much more complex than this last one, accounting not only the "problem" of an activity that contradicts rational principles and the normativity of reason (irrationality), but also the "problem" of "a failure in the process of the gradual development of a natural predisposition to the use of reason" (KLEIN, 2023, p. 101), as it will be further addressed. Therefore, irrationality is approached, here, from the perspective of society, i.e., as a harmful way of using reason that infiltrates social relations and institutions, and not merely as an individual phenomenon.

As it is clear, these clarifications show the intricacy of any attempt to show that a kind of *irrational* activity, such as rationalizing, can be used to address a source – let alone an important one – of *social irrationality*. The aim of the present paper is precisely to demonstrate the usefulness of Kant's concept of rationalizing for understanding the sources of social irrationality in our society. As it will become clear, Kant's definition and examples of rationalizing that we find throughout his works show us a harmful practice of human beings that not only occurs in an individual level, but also negatively affects the well-functioning and organization of social relations and institutions, which yields to social irrationality, according to the definition that will soon be addressed. To accomplish this tough task, I divide this paper in three sections, in addition to this introduction (section 1). Section 2 defines the concepts of irrationality and social irrationality according to Joel Klein's approach. The five main features of his account are briefly addressed. Section 3 brings Kant's definition of rationalizing as one can find in his works (3.1), as well as (3.2) the two main interpretations of this concept in Kant's works that one can find in the literature and, finally, (3.3) a synthesis of the main features that constitute Kant's approach to rationalizing. The final section (4), as expected, attempts to demonstrate the usefulness of Kant's concept of rationalizing for understanding the sources of social irrationality in our society.

[2] See Shel 2009, Guyer 2000, Papish 2018 and Sticker 2021.

2. Social irrationality

To address the concept of social irrationality from a Kantian perspective I will use Klein's (2023) approach, in which we find not only a clear definition of the concepts of social rationality and irrationality, but also a well-grounded explanation of how one can ground these notions in Kant's philosophy.

When answering the question whether a concept of social rationality is present in Kant's philosophy, or whether there could be, "in a Kantian perspective, a specific normative context of the use of reason that could be called social, and which would be distinct from the ethical, the juridical, the political, the epistemological, or the religious ones" (KLEIN, 2023, p. 100), Klein states that a "social normative context" is fundamental and ubiquitous to Kant's philosophy and "constitutes the internal link between the other normative contexts and principles [ethical, juridical, political, epistemological and religious]." (KLEIN, 2023, p. 100) Klein quotes a passage from the *Idea for a universal history with a cosmopolitan aim* (IaG) (IaG, AA 08: 18-19)[3] to corroborate the previous statement. The point Kant makes in the excerpt is that the social dimension constitutes the *medium* through which human beings' rational predispositions can develop and progress. The social dimension of human life enables the acquisition and transmission of *enlightenment* through generations. The following quotation from the *Anthropology from a pragmatic point of view* is related to this issue: "The human being is destined by his reason to live in a society with human beings and in it to *cultivate* himself, to *civilize* himself, and to *moralize* himself by means of the arts and sciences." (Anth, AA 07: 324-5) From these quotations, Klein states, we can identify three central characteristics of the social as related to the enlightenment:

> Firstly, it is related to the activity of people in a society. Secondly, the social element is rooted in a process of discovery (attempts and failures), learning (depuration, intelligibility, and retention of a correct use of a principle to produce a cognition, an action or a skill), and transmission (teaching and learning) of enlightenment. Finally, its aim must be the qualitative development of the different uses of reason and their reciprocal relations, which are promoted by the development of science and art (KLEIN, 2023, p. 101).

For the purposes of this paper, five aspects of Klein's approach must be highlighted:

First, Klein defines social rationality – from a Kantian perspective – as "the processual development of rational predispositions in society throughout several generations" (KLEIN, 2023, p. 101) and, consequently, *social irrationality* as "a failure in the process of the gradual development of a natural predisposition to the use of reason" (KLEIN, 2023, p. 101). In other words, considering that, for Kant, the development of the rational predispositions in the human beings is grounded in a *social* or *intersubjective* process "of discovery (attempts and failures), learning (depuration, intelligibility, and retention of a correct use of a principle to produce a cognition, an action or a skill),

[3] "*In the human being* (as the only rational creature on earth), *those predispositions whose goal is the use of his reason were to develop completely only in the species, but not in the individual.* Reason in a creature is a faculty of extending the rules and aims of the use of all its powers far beyond natural instinct, and it knows no boundaries to its projects. But reason itself does not operate instinctively, but rather needs attempts, practice and instruction in order gradually to progress from one stage of insight to another. Hence every human being would have to live exceedingly long in order to learn how he is to make a complete use of all his natural predispositions; or if nature has only set the term of his life as short (as has actually happened), then nature perhaps needs an immense series of generations, each of which transmits its enlightenment <*ihre Aufklärung*> to the next, in or der finally to propel its germs in our species to that stage of development which is completely suited to its aim. And this point in time must be, at least in the idea of the human being, the goal of his endeavors, because otherwise the natural predispositions would have to be regarded for the most part as in vain and purposeless" (*IaG*, AA 08: 18-19).

and transmission (teaching and learning) of enlightenment" (KLEIN, 2023, p. 101) throughout generations[4], one can also think, therefore, of a failure into which this social dynamics can fall.

Second, once social rationality concerns the conditions or the adequate means for the realization of rational predispositions in society, *the normativity that is proper to social rationality is connected to prudential and instrumental rationality* (KLEIN, 2023, p. 102)[5]. Moreover,

> *the normativity of social rationality deals* [...] *with public institutions* [...]. Social philosophy [...] engages with the normative, large-scale demands that are institutionalized in social practices. [...] It is the responsibility of social rationality [...] to organize social institutions (such as the family, the school, the workplace, associations, the press, and the internet) in such a way that they may also promote that end [, i. e. the development of rational predispositions] (KLEIN, 2023, p. 115. Added emphasis).

Third, *this "development of rational predispositions" can – and ought to – occur in different contexts* of the human being's life, seeing that rationality expresses itself normatively in different fields (e. g. in the epistemological as well as in the moral field). This way, one can find *social irrationality,* for instance, in the failure of "the [social] conditions for the gradual and constant development and transmission of epistemology and science", or even for "the proper gradual development of morality in society, i.e., *for those principles to have an increasingly correct and proper use in history*" (KLEIN, 2023, p. 101).

Fourth, *Klein distinguishes "failures and mistakes" from irrationality*, the former being the result of an improper use of a certain principle, the latter as "a *manner of thinking and acting* grounded on an erroneous principle" (KLEIN, 2023, p. 102), which brings the issue of rationality and irrationality to the methodological field of the legitimate and adequate principles which guide one's use of reason.

Fifth, Klein *identifies a source of irrationality*

> in the tension between animality and human rationality, or between the tendencies toward impulses and physical subjective conditions and the rational predispositions, as well as in the crude manner in which humans find a way to balance those different tendencies and predispositions in a systematic and moral unity (KLEIN, 2023, p. 103).

When not critically enlightened and disciplined, reason is subjugated by these impulses:

> It builds for itself principles that cannot be normatively valid since they have not been critically justified. This is how different types of egoism (logical, aesthetic, and moral, see Kant, Anth. AA 07: 128ff) or even the passions arise (see Kant, Anth. AA 07: 265ff.). In other words, certain impulses of our animality and sensibility tend to guide reason, which then becomes the servant of the passions by assuming and creating for itself strange and illegitimate principles. Social irrationality is this servitude and partiality of reason to animalistic nature and tendencies (KLEIN, 2023, p. 103).

As I will attempt to show, Kant's concept of *rationalizing* not only fits very well into the concept of *irrationality*, but, perhaps, under certain circumstances – namely, when it leads to a failure in

[4] "This process is closely related to the development of arts and sciences as Kant states in *Anthropology*: The sum total of pragmatic anthropology, in respect to the vocation of the human being and the characteristic of his formation, is the following. The human being is destined by his reason to live in a society with human beings and in it to cultivate himself, to civilize himself, and to moralize himself by means of the arts and sciences (Kant, Anth. AA 07: 324)" (KLEIN, 2013, p. 101).

[5] "After establishing the nature of the end, "the social" can be claimed to address the adequate means for the realization of that end, in other words, the process of the appropriate development of rational predispositions. The quest for the correct thing to do in order to achieve something is an aspect of practical philosophy. Rephrasing, the "how to do something" directed towards the means is a proper issue of hypothetical imperatives, both those of instrumental rationality (how to use things to achieve a certain end) and of prudential rationality (how to use other human beings to achieve a certain end), under a specific moral point of view.4 In other words, the proper normative feature of social rationality has to do with the appropriate means for the full development of natural predispositions of reason in the human species. *It is a matter of both instrumental and prudential rationality, which is subject to that specific moral end*" (KLEIN, 2023, p. 102).

the development of the rational predispositions of the human being throughout the generations, which not always is the case –, it can also be an important source of *social irrationality*.

3. Kant's concept of rationalizing <Vernünfteln>

3.1. Kant's definitions of *rationalizing*

The concept of *Vernünfteln*, often translated into English as rationalizing, was used by Kant in different works and refers to a "misuse" of reason by a rational being in which he engages in reflections that are grounded in an imperfect and unreasonable use of reason, creating an "air of truth" for something that is, in fact, illegitimate from a rational point of view. The main occasions in which Kant introduces a definition of *rationalizing* in his work are the following:

A) Although Kant used *Vernünfteln* and its correlates, in certain moments, in a non-pejorative sense (see KU V: 337n)[6], the term is most used by him with the already mentioned negative meaning, which is the case, for instance, of the *Anthropology from a Pragmatic Point of View* (Anth), in which the similar *Vernünftlei* – also translated as *rationalizing* – is defined as "a use of reason that misses its final end, partly from inability, partly from an inappropriate viewpoint" (Anth, AA 07: 200)[7], so that "reason is still different from rationalizing <*Vernünfteln*>, [which is] a playing with mere experiments in the use of reason without a law of reason." (Anth, AA 07: 228) Based on these statements, it could be said that one engages in *rationalizing* when one tries to justify certain claims based on false or illegitimate principles (see König, 2015, p. 2506). Still in the *Anth*, Kant states, regarding the reasonableness of believing in the existence of ghosts, that one "can *rationalize* about their possibility in all sorts of ways; but *reason* prohibits the *superstitious* assumption of their possibility, that is, without a principle of explanation of the phenomenon according to laws of experience" (Anth, AA 07: 228).

B) In the *Anthropology Busolt*, Kant states that "'Rationalizing' ought to mean 'using reason', but properly it is only an empty use of reason which contains nothing in regard to the true ends" (V-Anth/Busolt, AA 25: 1481).

C) The Groundwork of the Metaphysics of Morals (GMS) also provides us with a useful definition of rationalizing. At the end of the first section, after showing us that the principle of morality is already possessed by the "common human reason", Kant addresses the rhetorical question of whether this ordinary reason needs the help of philosophy in moral issues. The answer he gives us is that an innocent reason might need philosophy in order to acquire "access and durability for its precepts" (IV 4: 405), once this innocence, i. e. the "pre-reflective moral condition of common human reason" (Allison, 2011, p. 142), is "easily seduced" (GMS, AA 04: 405). Kant explains that this "seduction" is the result of something that lies in the human condition itself, namely the simple fact that the human being, due to his simultaneously rational and sensible nature, finds within himself a conflict between, on the one hand, the commands of duty – whose precepts are issued "unremittingly" <unnachlaßlich>– and, on the other hand, "the counterweight of his

[6] In the *Critique of the Power of Judgment* (KU) Kant states that a "rationalistic judgment" <*vernünftelndes Urteil*> is a necessary but insufficient condition for the faculty of judgment to fall into a dialectic; it merely means "any judgment that declares itself to be universal" (KU, AA 05: 337n). This way, "[a] power of judgment that is to be dialectical must first of all be rationalistic, i.e., its judgments must lay claim to universality, and indeed do so a priori, for the dialectic consists in the opposition of such judgments." (KU, AA 05: 336)

[7] On the same page, Kant seems to consider *Vernünftlei* and *räsonnieren* as interchangeable terms.

needs and inclinations, the entire satisfaction of which he sums up under the name happiness" (GMS, AA 04: 405). This gives rise to a "natural dialectic",

> that is, a propensity to rationalize <*vernünfteln*> against those strict law of duty and to cast doubt upon their validity, or at least upon their purity and strictness, and, where possible, to make them better suited to our wishes and inclinations, that is, to corrupt them at their basis and to destroy all their dignity" (GMS, AA 04: 405)

In short, "to rationalize" is understood by Kant, in the moral domain, as the attempt to find pseudo-justifications that could give the appearance of an agreement between the two opposing demands previously mentioned, which is made through an (obviously) illegitimate modification of the so-called "law of duty". The propensity of human beings to rationalize is what constitutes what Kant called a "natural dialectic". It is "dialectical" because it corresponds to a tension between happiness and morality as two possible determining grounds of an agent's will (see Sticker, 2021, p. 16). It is "natural" because this tension originates in the dual nature of the human being, as rational and sensible.

3.2. A (brief) "state of the art"

A short look at the state of the art of Kant's concept of *rationalizing* can help us to further clarify this issue. In the last decades, there has been an increase in approaches to this topic. While the approaches of Shell (2009) and Guyer (2000) addressed this topic briefly, authors such as Laura Papish (2018) and Martin Sticker (2021) provided more detailed approaches. Papish's and Sticker's approaches are especially useful for the purposes of this paper and will, therefore, be addressed.

3.2.1 Laura Papish

Although Papish is attempting to show us the role of the human being's rationalizations for the phenomenon of moral self-deception, Papish's account seems to be broad enough to encompass not only moral *rationalizing* but also *rationalizing* in theoretical matters. After stating that rationalizing, for Kant, means a "misappropriation of our cognitive powers to the detriment of reason's true or final ends" (PAPISH, 2018, p. 73-4), as we find in the already mentioned passage of *Anthropology Busolt*, Papish says that, following some of Kant's statements in Wiener and Dohna Logic, "it seems that he describes the true end of reason as wisdom in the logic lectures because 'wisdom' [*Weisheit*] can account for how reason has both theoretical aims such as truth and practical aims concerning morality and prudence" (PAPISH, 2018, p. 78n). In addition, Papish writes that Kant's concept of *rationalizing* presupposes the introduction of a "desirable cognition" or "hoped for justification" into the reasoning process (PAPISH, 2018, p. 74), which evidently contradicts reason's end of wisdom, and that rationalizing "involves a shift in attention to a more attractive and true yet comparatively less relevant alternative cognition" (PAPISH, 2018, p. 75). All these characteristics that Papish attributes to rationalizing allow its application to both moral and theoretical fields, even though it highlights the fact that this harmful activity always has its root in a moral issue.

The three examples of *rationalizations* that Papish finds in Kant's works are also useful for the present approach. The first one is found in *Toward Perpetual Peace,* in the appendix *On the disagreement between morals and politics with a view to perpetual peace*, when Kant discusses about the meaning of a legitimate political practice and argues against the so called "political moralist", opposing it to the "moral politician". While the last "takes the principles of political prudence in such a way that they can coexist with morals" (ZeF, AA 08: 372), the former "frames a morals to

suit the statesman's advantage" (ZeF, AA 08: 372). While the last, in problems of practical reason, begins from a formal and unconditional practical principle, the former begins from a material one (ZeF, AA 08: 376-7), wrongly trying to derive from experience or the "mechanism of nature" the principles of right and politics. That's why, Kant says, the principle of the political moralist is merely a "technical problem", which takes perpetual peace as a matter of "natural good", whereas that of the moral politician is a moral one (ZeF, AA 08: 377). Finally, Kant states, while it is a principle of moral politics "that a people is to unite itself into a state in accordance with freedom and equality as the sole concepts of right" (ZeF, AA 08: 378), which is based upon duty instead of prudence, the political moralists "reason subtly" <*vernünfteln*>, i. e. they rationalize

> about how the natural mechanism of a multitude of human beings entering into society would invalidate those principles and thwart their purpose, and also try to prove their contention against them by examples of badly organized constitutions of ancient and modem times (e.g., of democracies without a representative system) (ZeF, AA 08: 378).

The political moralists, Papish stresses, "introduce a new and, to them, more desirable cognition that respects the truth of the original cognition concerning our internal, *a priori* moral principles of public right while refusing to stay focused on that truth" (PAPISH, 2018, p. 74).

The second example Papish provides us concerns Kant's argument, in the *Doctrine of Right* (MS, AA 06: 318), against those who engage in historical inquiries about how their sovereign came to power. These, Kant says, are "pointless subtle reasonings <*Vernünfteleien*>", considering "this genetic point does not contradict arguments about the sovereign's right to rule" (PAPISH, 2018, p. 74).

Papish's third and last example concerns Kant's essay *What is Enlightenment?*, in which the philosopher of Königsberg states that an officer who, when receiving an order from his superiors, decides, while on duty, to engage in subtle reasoning <*vernünfteln*>, i. e. to rationalize about its appropriateness <*Zweckmäßigkeit*> and utility (WA, AA 08: 37). As Papish writes, Kant

> is quite clear that these officers do not directly assert that their superiors hold their position unlawfully. This would be a lie that directly contradicts what the officers know to be true, so instead the officers pose questions about the demands issued by these authorities in a way that only a private citizen should (WA 8:37). As with Kant's other examples, the rationalization in question involves a shift in attention to a more attractive and true yet comparatively less relevant alternative cognition (PAPISH, 2018, p. 74-5).

3.2.2. Martin Sticker

Two important differences between Papish's and Sticker's accounts must be highlighted. *First,* while the former denies that *rationalizing* can lead the subject to an absolute (false) certainty in his incorrect reasoning, assuming that rationalizers are able to recognize their need for help, the later, on the other hand, although recognizing the virtues of this view insofar as it stresses that the rationalizer always grasps, to some extent, the irrationality of his reasoning, defends that the last perspective underestimates the dangers of the impact of *rationalizing* in one's capacity to correctly reflect on moral matters (STICKER, 2021, p. 39-40). *Second,* while Sticker, rather than making a broad approach that encompasses both theoretical and practical spheres, focuses his account of Kant's concept of *rationalizing* on the moral sphere, i. e. on its effect on the agent's grasp of morality, as well as on the necessary conditions for an agent to rationalize against moral

commands[8], Papish's account, although without explicitly stating that, seems to encompass both theoretical and practical rationalizing.

Despite these differences with respect to Papish's account, the most important point for us is that Sticker seems to understand *rationalizing*, at least in the moral sphere, in a similar way as Papish but focusing on "different paradigmatic cases of the same phenomenon" (STICKER, 2021, p. 8). While Papish helps us to find everyday cases of *rationalizing* in the political and social sphere, Sticker focuses on Kant's extreme examples of *rationalizing*, namely his discussions of eudaemonistic theories and religious practices. The complementarity of both accounts and, therefore, their usefulness for an understanding of the meaning of the harmful phenomenon of *rationalizing* within Kant's philosophy will become evident.

In the following, I will briefly reconstruct two elements of Sticker's approach which will be useful for our understanding of the meaning of the harmful phenomenon of *rationalizing* within Kant's philosophy, as well as for its application to contemporary problems of social irrationality, namely Sticker's analysis of *rationalizing* as a moral corruption and the connection he makes between the concepts of *rationalizing* and *ideology*.

Concerning the first element, Sticker focuses his analysis of *rationalizing* on the way Kant addresses the term at the end of the already briefly discussed first section of the GMS. This way, Sticker considers that "[r]ationalizing ultimately amounts to challenging the 'validity' [10] of the moral law" (STICKER, 2021, p. 16-7), which, as he states, does not lead to the conclusions that the rationalizer renounces his commitment to morality. On the contrary, the rationalizer rationalizes precisely with the aim of reconciling two contradictory although (for him) important demands: the rational and the sensible one. The rationalizer does not accept either the pain of giving up his sensible desire or the pangs of conscience for not being rationally justified. The interest in being rationally justified, Sticker states, "is rooted in an agent's acknowledgement of the authority of duty and pushes them to devise excuses and pseudo-justifications" (STICKER, 2021, p. 30). With this purpose, the rationalizer modifies his conception of morality into a "better one", which precisely means the *corruption* of the moral law mentioned by Kant (GMS, AA 04: 405). Sticker considers that this *rationalizing* is divided into two possible strategies, namely questioning the *purity* and the *strictness* of duty. As one can easily anticipate, the former concerns casting doubt on the idea that "nothing empirical [...] functions as criteria for moral evaluation" (STICKER, 2021, p. 18) and that obligatory actions must be motivated by respect for the moral law; the later concerns the idea that (perfect) duties never admit of exceptions.

Concerning the second elements, it is useful for this work the connection Sticker identifies between these "apparent justifications" that characterize *rationalizing*, on the one hand, and the concepts of *ideology* and *uncritical philosophy*, on the other hand. When addressing the possibility of an agent being deceived or confused due to his *rationalizing*, Sticker asserts that although, for Kant, an agent can never think that his moral transgressions are fully justified, they can find "subjective reasons" (V-MS/Vigil, AA 27: 617) or "subjective grounds of consolation" (V-MS/Vigil, AA 27: 618-9) for these violations. These reasons are subjective insofar as they lack the

[8] "'Rationalizing', as I will use the term, is self-deception about *moral* matters. There are other kinds of self-deception, such as about what is prudent, one's capabilities and social standing and maybe even about purely theoretical questions. Self-deception about these issues, however, requires a framework different from the one that explains rationalizing in my sense, since these other forms of self-deception are not driven by a rational interest in being morally justified" (STICKER, 2021, p. 6).

objective force of those reasons that *ought to* be accepted by others and by oneself when impartially analyzed (STICKER, 2021, p. 38). In other words, they are merely "apparent justifications" or

> psychological means to cope with one's failure to live up to the demands of the moral law. They share important structural properties with genuine justifications; they are meant to apply to all agents in relevantly similar circumstances, and they cover a multitude of similar cases. [...] Apparent justifications therefore do not merely seemingly sanction one-off transgressions, but potentially condone systematic violations of the moral law (STICKER, 2021, p. 38).

This "systematicity" of pseudo-justifications, Sticker recalls us, is expressed in Andrews Reath's statement that *rationalizing* or "the influence of self-love on the will is sustained by an ideology of sorts, which enables individuals to view their maxims as objectively acceptable reasons." (REATH, 2006, p. 21) Sticker understands the ideological aspect of *rationalizing* as its capacity to form "a system of beliefs" or a set or propositions that support each other, which is precisely what makes it more difficult for the rationalizer to be corrected or challenged in his false beliefs. Nevertheless, Sticker still recognizes that the kind of "certainty" that the rationalizer has need to be different from the kind of certainty that justifiable reasons can provide:

> Certainty can also be merely subjective, namely the kind of unwarranted certainty that a fallible and imperfect agent might have in a belief even though this belief is, in fact, unjustified. Certainty of the latter kind can be the result of rationalizing or of other mistakes in reasoning, and such a certainty can be psychologically powerful, action-guiding and difficult to overcome (STICKER, 2021, p. 40).

It is difficult to state, from a Kantian point of view, what degree of criticism a rationalizer deserves, once he seems to be partly deceived and partly conscient that there is some problem in his reasoning – especially in moral matters, in which the grasping of one's duties is easily carried out by the agent.

To finish this subsection, I bring the two paradigmatic examples of the extreme cases of *rationalizing* and *corruption* in the moral field that Sticker addresses. The first example concerns the "kind of rationalizing" that leads the agents to "eudaemonistic forms" of morality. Sticker recalls us the of Kant's critics of Christian Garve's moral theory in the essay *Theory and Practice* (TP, AA 08: 284-5). Garve would have confused the moral concept of duty and the human being's pursuit for happiness, once he cannot conceive other motivation for actions than this one: "[h]e has adopted the wrong metaphysical framework to understand the possibility of acting from duty" (STICKER, 2021, p. 41). The second of Sticker's examples concerns the idea that religious practice can also be an important source of ideology: it "creates supposed grounds of excuses, apparent justifications and means to seemingly escape one's responsibility" (STICKER, 2021, p. 41-2) To illustrate that, Sticker mentions Kant's "inquisitor case", which regards an individual that, due to his religious faith, decides to take someone's life (RGV, AA 06: 186).

Despite the richness of Sticker's account, it must be stated that Kant's concept of *rationalizing* can be broadly addressed – similarly as Papish did –, in a way that comprises not only its manifestation in the moral field as a harmful or illegitimate kind of moral reflection, but also in the theoretical sphere. In other words, if we consider the already addressed general definitions of the concept of *rationalizing* provided by Kant (see section 3.1 and 3.2.1), we cannot fail to define *rationalizing* as a type of prejudicial use of reason that can be done, in principle, in all the spheres that reason introduces itself normatively through the establishing of the principles for a correct or objective judgment. Kant's broad definitions of *rationalizing* (Anth, AA 07: 200; V-Anth/Busolt, AA 25: 1481), as well as his already mentioned amusing example of the *rationalizing* about the existence of ghosts through a dismissal of a "[rationally required] principle of explanation of the phenomenon

according to laws of experience" (Anth, AA 07: 228), show us the scope of his application of the concept.

3.3. Essential aspects of rationalizing

In order to make Kant's concept of rationalizing clear and emphasize all its essential characteristics, five points must be highlighted:

A) Concerning a presupposition for the concept of *rationalizing*, a *rationalizing* subject is always *making use* of reason – even though an illegitimate or sophistical one – or is, to some extent, committed to reason. Although this might sound like a truism, it still is important to highlight that the *rationalizing* activity is necessarily made by rational beings, once it is characterized precisely by a use of reason against itself. One could state that rationalizing is, to some extent, a non-critical use of reason. Similar to this statement is the idea that *irrationality* or *contradiction* is something that cannot be carried out by non-rational beings, once only *non-rationality*, instead of *irrationality,* comes from them.

B) In Kant's works, *rationalizing* seems to have a close relation with the concepts of dialectic and illusion, once the "misuse" of reason that characterizes *rationalizing*, if not always, at least has a tendency to lead the subject to have a false or distorted understanding of the demands of reason – however, this relation between *rationalizing* and dialectic or illusion is not always indicated by Kant in the moments he addresses the *rationalizing*[9].

C) As already addressed at the end of the last subsection, Kant thought of *rationalizing* as a prejudicial use of reason that can occur both on the practical and the theoretical spheres, i.e. in all the spheres in which our reason has a normative force.

D) *Rationalizing* is usually taken by Kant to be an *individual* activity, i.e. a misuse of reason through the construction of illegitimate principles carried out by a rational subject in his personal reasoning on practical or theoretical matters. Kant's example of the public officer whose *rationalizing* is directed to his superiors in order to challenge their authority (WA, AA 08: 37) seems to be the only case of *rationalizing* illustrated by Kant that explicitly involves a dialogical scenario. Nevertheless, one can still think of the possibility of *rationalizing* activity being carried out publicly without exceeding the limits of Kant's philosophy, namely when one person attempts to convince another through rationalizations.

E) *Rationalizing* seems to always involve a moral problem: the choice to disrespect reason's normativity and the limits established by its critic with the aim of satisfying some personal desire. Therefore, one cannot rationalize without having some level of responsibility for that.

4. Rationalizing and social irrationality

As already addressed in the second section of the present paper, the concept of *social* irrationality is more complex than that of irrationality *in general*. The reason for that lies on the fact that, whereas the latter denotes the rationally illegitimate activity of adopting inappropriate or false principles, the former, in its definition, refers to a failure in the gradual development of the rational predisposition of the human being, which encompasses the irrational activity of adopting illegitimate principles

[9] While in the mentioned passages of the *Anth* the concept of rationalizing is not connected to that of dialectic or illusion, in the GMS this relation is unequivocal (GMS IV: 405).

but is not limited to it. The "social" aspect of both concepts of *social rationality* and *social irrationality* states that, in dealing with these concepts, one is addressing the social or intersubjective conditions for the development of the rational predispositions throughout generations.

On the one hand, the adequacy of the concept of *rationalizing* to the concept of *irrationality* addressed in the second section is easily justifiable. On the other hand, to address the adequacy of the former concept to that of social *irrationality* is much more complicated, once the latter involves more elements.

Concerning the adequacy of *rationalizing* with the concept of *irrationality,* as one can see from the definition and examples provided, Kant's concept of *rationalizing* concerns an illegitimate use of reason – i. e. one that contradicts its end, namely, wisdom – or its misappropriation through the inadequate introduction of a "desired cognition" or "hoped for justification" in the reasoning process. If Kant's definition seems to lack sufficient clarification, it can be further elucidated from the already mentioned examples of *rationalizing*. There, one can note that Kant is addressing an appropriation of reason which is essentially characterized by the call for a rationally rejectable principle or "way or reasoning" with the aim of constructing an appearance of truth and reasonability for what is not. Therefore, the problem of the *rationalizer* is not of a misapplication of a legitimate principle, but of a creation of false principles of reflection, which can be made in different contexts of rationality. Moreover, *rationalizing* encompasses both theoretical and practical fields – those in which reason imposes normativity. This, along with the just addressed definition of *rationalizing*, attests its fitness to the concept of irrationality, which can be not only practical, but also theoretical irrationality.

Regarding the relation between *rationalizing* and *social* irrationality, it depends upon the fact that the phenomenon of *rationalizing*, in the different contexts and ways in which it manifests itself, disturbs the development of the human being's rational capacities, i. e. puts hindrances in the process of achieving and/or transmitting a degree of enlightenment to future generations. To address the possibility of such a phenomenon, it might be useful to introduce a scheme that differentiates contexts and ways of *rationalizing*. With such scheme, we will take a step in the direction of understanding the way *rationalizing* can spread in society and, therefore, the potential harm such activity to the development of rational predispositions through generations. In other words, we will be closer to comprehending how *rationalizing* can affect the social spaces and institutions that are important for the development of rational predispositions, which leads to a problem of social irrationality.

The construction of the mentioned scheme will take into consideration (a) a distinction in the object of *rationalizing* and (b) a distinction in the actors of *rationalizing*. Regarding (a), such distinction occurs between *rationalizing* in the (a1) moral and in the (a2) theoretical sphere – i.e. regarding moral and theoretical matters or objects. Regarding (b), such distinction is made between *rationalizing* as a (b1) private, a (b2) public and an (b3) intersubjective activity. With a (b1) private form of *rationalizing* it is understood the *rationalizing* that is carried out alone by the individual in his solitary reflections. With (b2) a public form of *rationalizing* it is understood the *rationalizing* that is made by one or more *rationalizing* actors over one or more *non-rationalizing* deceived or misled agents. Finally, with a (b3) intersubjective form of *rationalizing* it is understood the *rationalizing* that is shared between two or more *rationalizing* actors, i. e. a *rationalizing* reflection or reasoning that is made communally. This way, one can rationalize:

(b1) *privately* or "to oneself"

(b1a1) in *moral* matters. This is the case when the subject carries out a private *rationalizing* reflection on moral matters, regarding either his past or future choices, with the aim of constructing a form of morality better suited for his subjective desires. Examples of this in Kant's works are Kant's already mentioned passage of the "natural dialectic" in the GMS (4: 405) and his understanding, also already mentioned and present in the *Doctrine of Right*, that "a subject *ought not to rationalize* for the sake of action [*werktätig vernünfteln*]" (MS, AA 06: 318) about the origin of "the supreme authority to which it i subject" (MS, AA 06: 318). Kant's concern, here, is not exactly with publicizing such rationalizations or using it to argue with others, but with the subject that, "having pondered over the ultimate origin of the authority now ruling, wanted to resist this authority" (MS, AA 06: 318); in other words, he is discussing about private instead of public or intersubjective rationalizing. Nonetheless, of course, one could also think of someone rationalizing about the same issue to another instead of to oneself.

(b1a2) in *theoretical* matters, which means carrying out a private rationalizing reflection on theoretical or non-moral matters with the aim of providing justifications for the unjustifiable according to principles of reason. Kant's already mentioned example of rationalizing about the existence of ghosts seems to be an example of this (Anth, AA 07: 228).

(b2) *publicly* or "to others"

(b2a1) in *moral* matters, i.e. when one attempts to deceive someone else – regarding moral matters – through a *rationalizing* strategy of argumentation, leading the other to a misleading moral thinking in order to satisfy some personal interest. Kant's already mentioned example of the public officer seems to be the illustration of an *attempt* of such a way of *rationalizing* (WA, AA 08: 37). Moreover, Kant's also already mentioned example of the "political moralist" perhaps can also fit here. The "political moralist" provides illegitimate or false principles through which he pretends to seemingly ground a conception of the state and of public institutions, i. e. of politics that is better fitted to his personal interests (ZeF, AA 08: 372, 376-8). However, this nefarious figure does not aim to deceitfully prove these false ideas to himself, as in the case of a private way rationalizing, but to the others, in order to infect their moral convictions and bring them to agree with him in his pursuit of his personal interests.

(b2a2) in *theoretical* matters, i.e. when one attempts to deceive someone else – regarding theoretical matters – through a *rationalizing* strategy of argumentation, leading the other to a misleading conception about a theoretical matter.

and (b3) *intersubjectively* or "with others"

(b3a1) in *moral* matters. One rationalizes intersubjectively in moral matters when two or more persons have a common rationalized conception and/or engage in a shared *rationalizing* reflection on moral matters. Through this form of *rationalizing* one can find in the other support for his own rationalizations. For instance, when a rationalizing subject finds himself inside a bubble constituted by persons that share the same *rationalizing* thinking and, in this way, do not keep in touch with conceptions that deconstruct their *rationalizing* certainties.

(b3a2) in *theoretical* matters, when two or more persons have a common rationalized conception and/or engage in a shared *rationalizing* reflection on theoretical matters. As in the intersubjective *rationalizing* in moral matters, one can find here a strong support for his own rationalizing.

The previous scheme makes it clear the scope of the harmful effect of rationalizing. However, until now rationalizing seems to be merely an irrationality in both moral (ethical and juridical) and theoretical matters, either privately, publicly or intersubjectively, but not specifically concerning social irrationality, at least not according to the definition already addressed.

Let us give one step behind and briefly recall a crucial feature of social rationality already addressed. As seen, Klein states that social rationality concerns "the processual development of rational predispositions in society throughout several generations" (KLEIN, 2023, p. 101), while the "social" concerns the appropriate means for the realization of this end. Thus,

> [i]t is the responsibility of social rationality [...] to organize social institutions (such as the family, the school, the workplace, associations, the press, and the internet) in such a way that they may also promote that end [, i. e. the development of rational predispositions] (KLEIN, 2023, p. 115).

Consequently, "the normativity of social rationality deals [...] with public institutions" (KLEIN, 2023, p. 115), and social irrationality, as it is more than clear at this point, concerns a hindrance or failure in this organization and, therefore, in the mentioned process.

Let us now turn back to the concept of rationalizing as it is addressed by Kant in his different works. Let us also consider the scheme previously constructed, which helped us in understanding the different contexts and ways in which Kant's rationalizing might occur. It seems to be the case that the harmful activity of rationalizing is not only detrimental from an individual point of view, but also from a social point of view, i.e. from the point of view of the well-functioning of social institutions. To state that rationalizing is individually or subjectively harmful merely means that the irrationality that constitutes this activity affects the way the rationalizer, as an individual, reason about a certain matter. On the other hand, to affirm that rationalizing is harmful for a certain social institution means that it is no longer – if it ever was – or, at least, it is less oriented to promote the development of the rational predispositions of the individuals that pertain to the community in which this institution operates. Considering that rationalizing is characterized by the creation of an apparently rational reasoning based on illegitimate or false principles with the aim of reaching a personal end, there is no reason not to think on the social institutions as potential victims of this harmful way of reasoning, once these same institutions are subject to the correct or incorrect, benefic and detrimental leading of individuals. The appropriate organization of social institutions aiming for the promotion of the development of rational predispositions is not something that belongs to the nature of the functioning of these institutions. On the contrary, an institution may very well be used to promote irrationality.

Kant's examples of rationalizing might once again be useful. While the "officer case" illustrates a misuse of one's capacity to criticize an authority through a rationalizing argumentation that corrupts the workspace, the "political moralist case", which is even more expressive of what is being dealt with here, instances a corruption of political social institutions by those who try to submit them to their personal interests, which in this case is not done through coercion as in a seizure of power, but through a deceitful argumentation and the construction of spurious principles that make their political thinking more adequate to their personal ends. If the political moralist comes to control the state, i. e to organize and orient its functioning, this will hardly agree

with the rational organization of a social institution aiming for the development of its member's rational predispositions.

5. Concluding Remarks

The present paper argued that Kant's concept of *rationalizing* <*Vernünfteln*> can be understood as a conceptual tool to address the phenomenon of social irrationality in our society. After clarifying how one can address the concepts of *irrationality* and of *social irrationality* within Kant's philosophy, as well as Kant's conception of *rationalizing* and its inherent relationship with the notion of irrationality, the paper defended that we could use the concept of rationalizing to comprehend the way social irrationality manifests in our society.

Social irrationality, as addressed, concerns "a failure in the process of the gradual development of a natural predisposition to the use of reason" (KLEIN, 2023, p. 101), so one can find social irrationality in the failure of "the [social] conditions for the gradual and constant development and transmission of epistemology and science" and for "the proper gradual development of morality in society, i.e., for those principles to have an increasingly correct and proper use in history" (KLEIN, 2023, p. 101). Moreover, as seen, irrationality concerns "a manner of thinking and acting grounded on an erroneous principle" (KLEIN, 2023, p. 101). On the other hand, Kant's concept of rationalizing is defined as a rational being's "misuse" of reason in which one engages in moral or theoretical reflections grounded in an illegitimate or non-critical use of reason, possibly leading him to have a false or distorted understanding of the demands of reason. Finally, considering the different contexts and ways of rationalizing addressed and its capacity to affect the social spaces and institutions that are relevant for the development of rational predispositions, it becomes clear how rationalizing can spread in society and hinder the development of these predispositions through generations, leading to social irrationality.

Bibliographic References

ALLISON, H. 2011. *Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Oxford: Oxford University Press.

GUYER, P. 2000. *Kant on Freedom, Law, and Happiness*. Cambridge: Cambridge University Press.

KANT, I. 1996a. An answer to the question: What is Enlightenment? [*WA*]. GREGOR, M. (Ed.) *Practical Philosophy*. Cambridge: Cambridge University Press.

KANT, I. 1996b. The Metaphysics of Morals [*MS*]. GREGOR, M. (Ed.) *Practical Philosophy*. Cambridge: Cambridge University Press.

KANT, I. 1996c. Toward perpetual peace [*ZeF*]. GREGOR, M. (Ed.) *Practical Philosophy*. Cambridge: Cambridge University Press.

KANT, I. 1997a. G*roundwork of the Metaphysics of Morals* [GMS]. GREGOR, M. (Ed.) Cambridge: Cambridge University Press.

KANT, I. 1997b. *Lectures on Ethics*. HEATH, P.; SCHNEEWIND, J. B. (Eds.) Cambridge: Cambridge University Press.

KANT, I. 2000. *Critique of the Power of Judgment* [KU]. GUYER, P. (Ed.) Cambridge: Cambridge University Press.

KANT, I. 2007. Anthropology from a pragmatic point of view [Anth]. In: ZÖLLER, G.; LOUDEN, R. *Anthropology, History, and Education*. Cambridge: Cambridge University Press.

KANT, I. 2012. *Lectures on Anthropology*. WOOD, A. W.; LOUDEN, R. B. (Eds.). Cambridge: Cambridge University Press.

KLEIN, J. T. 2023. Enlightenment as the normative principle of social rationality. In: *Studia Kantiana*, vol. 21, n. 1, 99-117.

KÖNIG, P. 2015. Vernünfteln. In *Kant-Lexicon*, eds. Marcus Willaschek, Jürgen Stolzenberg, Georg Mohr, Stefano Bacin, 969-973. Berlin/Boston: Walter de Gruyter.

PAPISH, L. 2018. *Kant on Evil, Self-Deception, and Moral Reform*. Oxford: Oxford University Press.

SHELL, S. 2009. *Kant and the Limits of Autonomy*. Cambridge, Mass.: Harvard University Press.

STICKER, M. 2017. When the Reflective Watch-Dog Barks Conscience and Self-Deception in Kant. In: *Journal of Value Inquiry*, 51(1), 85-104.

STICKER. 2021. *Rationalizing (Vernünfteln)*. Elements in the Philosophy of Immanuel Kant. Cambridge: Cambridge University Press.

Toward a Kantian theory of prudential irrationality: between intellectual error and volitional failure

Rumo a uma teoria kantiana da irracionalidade prudencial: entre o erro intelectual e a falha volitiva

Tales Yamamoto[1]
Universidade Federal de Santa Catarina (UFSC)/Universität Vechta
talesyamamoto@hotmail.com

Abstract: This article investigates the conditions for a Kantian theory of prudential irrationality. Against Merle (2023), it argues that intellectual errors, such as incorrect beliefs, do not suffice to generate irrational actions. The analysis focuses instead on whether volitional failures can lead to prudentially irrational actions. To examine this, two interpretive models are considered: the negative model, inspired by Timmerman (2022), which denies prudential irrationality, and the positive model, developed by Korsgaard (2008), which affirms it. While Timmerman strips the Hypothetical Imperative of normativity, Korsgaard subordinates it to the Categorical Imperative. Both models face limits: the first excludes the possibility of instrumental irrationality; the second risks expanding the moral domain or weakening the link between freedom and the moral law. The challenge remains to explain how instrumental rationality is possible without reducing it to morality, while preserving its intrinsic tie to freedom.

Keywords: categorical imperative; hypothetical imperative; instrumental reason; Kant; normative ethics; rational agency.

Resumo: Este artigo investiga as condições para uma teoria kantiana da irracionalidade prudencial. Contra Merle (2023), defende-se que erros intelectivos, como crenças incorretas, não são suficientes para gerar ações irracionais. A análise concentra-se, em vez disso, na possibilidade de falhas volitivas conduzirem a ações irracionais prudenciais. Para examinar essa hipótese, consideram-se dois modelos interpretativos: o modelo negativo, inspirado em Timmerman (2022), que nega a irracionalidade prudencial, e o modelo positivo, desenvolvido a partir de Korsgaard (2008), que a afirma. Enquanto Timmerman esvazia o Imperativo Hipotético de normatividade, Korsgaard o subordina ao Imperativo Categórico. Ambos os modelos enfrentam limites: o primeiro exclui a possibilidade de irracionalidade instrumental; o segundo arrisca ampliar excessivamente o domínio moral ou enfraquecer o vínculo entre liberdade e lei moral. O desafio permanece em explicar como a racionalidade instrumental é possível sem reduzi-la à moralidade, preservando ao mesmo tempo sua ligação intrínseca com a liberdade.

Palavras-chave: imperativo categórico; imperativo hipotético; razão instrumental; Kant; ética normativa; agência racional.

[1] I had the privilege of discussing earlier versions of this paper with many colleagues. A thank you to Frank Rettweiler, Gehad Marcon Bark, Luciana Martinez, Marina Guimarães Back, Nicole Martinazzo and Sulamith Weber. Very substantial contributions were offered by some colleagues and professors. A very special thank you to Cristina Foroni Consani, Eduardo de Oliveira da Costa, Eduardo Estevão Quirino, Jean-Christophe Merle and Joel Thiago Klein. I would also like to thank the two anonymous reviewers for their contributions. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and is part of the PROBRAL Project "Kantian Perspectives on Social Irrationality" (CAPES-DAAD), process number 88887.078809/2024-00.

Recebido em 31 de julho de 2025. Aceito em 23 de novembro de 2025.

1. Introduction

If we want to ask *how* an irrational action is possible, firstly we must ask *whether* irrational actions are possible. This skepticism is valid insofar as Kant refers multiple times in the *Groundwork* (therefore GMS) and in the *Critique of Practical Reason* (therefore KpV)[2] to an action, but not to an irrational action. More specifically, in this case, we want to ask whether *prudential* irrational actions are possible. The question about prudential irrational actions is, I believe, a bit more tricky than the question about moral irrational actions. If we want to take Kant's philosophy and create a theory about moral irrationality, I see that as an easy path to take. In the case of prudential irrationality, I do not think that it would be that simple.

My aim in this paper is modest: to lay out some conditions for a possible Kantian model of prudential irrationality. The problem, then, is to understand how we can ideally make sense of all the relevant parts of Kant's work in order to address this issue. In this reconstruction, my main concern is not to offer an account of how Kant might be used to philosophically engage with contemporary problems of irrationality. Rather, I wish to sketch a model of how Kant himself would have thought about irrationality. The former type of work could freely abandon, or at least set aside, some very important theses and arguments of Kant's philosophical system. In this paper, however, I will attempt to make sense of Kant's (hypothetical) position on irrationality while considering all the relevant aspects of his philosophy. This entails that the most accurate reading of Kant's work may lead us to conclude that his account of this topic is not suitable for today's debates on irrationality.

For methodological clarity, I provide in section two an account of the meaning of certain concepts central to the discussion, such as 'rationality,' 'irrationality', 'action,' and 'prudential action'. In section three, I argue that irrational actions can only be conceived in opposition to practical rationality, and that they cannot be understood as mere products of intellectual mistakes or errors. Finally, in section four – which I take to be the central part of the paper – I compare current interpretative models of Kant's account of prudential irrationality, highlighting both their advantages and, especially, their limits. My conclusion will be that the two predominant models fail to provide an account that both preserves key elements of Kant's philosophy and allows us to further elaborate a theory of instrumental irrationality.

2. Some initial conceptual presentations

2.1. Rationality

Rationality has at least two central meanings in Kant's work: the rationality derived from (i) the good use of understanding and the one derived from (ii) the practical use of reason. In this sense, being theoretically rational means being able to make a good use of the (general) faculty

[2] In this paper, I adopt conventional models of citation and employ a series of abbreviations. The *Critique of Pure Reason* (KrV) is cited according to the A/B pagination. All other works of Kant are referenced following the *Akademie-Ausgabe* (AA) convention: *Critique of Practical Reason* (KpV, AA 05); *Critique of the Power of Judgment* (KU, AA 05); *Groundwork of the Metaphysics of Morals* (GMS, AA 04); *The Metaphysics of Morals* (MS, AA 06); *What Does It Mean to Orient Oneself in Thinking?* (WDO, AA 08); and *An Answer to the Question: What Is Enlightenment?* (WA, AA 08). For the sake of readability, I also abbreviate Categorical Imperative as CI, Hypothetical Imperative as HI, and hypothetical imperatives as HIs.

of knowledge (KU AA 05: 174-176, 195)[3]. Therefore, in its theoretical aspect, a rational agent is one that is constantly using his general faculty of knowledge in a critical way. A rational agent is also one that is not in a position of cowardice of laziness and can use his own understanding without the guidance of others (WA, AA 08: 35). In other words, he is one capable of thinking for himself[4] (WDO, AA 08: 146)[5]. Now let us consider practical rationality.

In the KpV, Kant argues for the practical reality of the freedom (KpV AA 05: 3-4). The argument is possible only because Kant developed a very distinct framework to deal with the relation with the object. While the general faculty of knowledge only seeks to represent an object, the faculty of desire, beyond the representation of the form of the a priori moral law, also seeks to create the object of desire through this representation (KpV AA 05: 44-45; KU AA 05: 178). This central distinction implies also a central differentiation between the practical rationality and the theoretical one. While the latter wants to represent correctly the object (i.e., according to the good use of the faculty of knowledge), the former seeks to create it[6].

For now, let us follow the *Groundwork* and grant that there are two general practical principles: one in which the desired end is conditional (the Hypothetical Imperative – hence HI), and another in which the desired end is unconditional (the Categorical Imperative – hence CI). Now, let us *assume, only provisionally*, that a rational agent, in their *practical realm*, is one that *(i) recognize his/hers desired ends* (whether conditional or unconditional) (*intellectual aspect*)[7], and *(ii) is capable of acting as a cause* in bringing about the objects in accordance with the representation of those desired ends (*volitional aspect*)[8].

[3] In a more strict sense, it also means being disciplined against the temptation of taking the ideas of reason as having a constitutive theoretical reality (B 421, B 737-740), and, therefore, not to take as objective what is merely subjective (KrV B 353–354).

[4] "Thinking for oneself means seeking the supreme touchstone of truth in oneself (i.e., in one's own reason); and the maxim of always thinking for oneself is enlightenment. Now there is less to this than people imagine when they place enlightenment in the acquisition of information <*Kenntnisse*>, for it is rather a negative principle in the use of one's faculty of cognition, and often he who is richest in information is least enlightened in the use he makes of it. To make use of one's own reason means no more than to ask oneself, whenever one is supposed to assume something, whether one could find it feasible to make the ground or the rule on which one assumes it into a universal principle of the use of reason. This test is one that everyone can apply to himself; and with this examination he will see superstition and enthusiasm disappear, even if he falls far short of having the information to refute them on objective grounds. For he is using merely the maxim of reason's self-preservation" (WDO, AA 08: 146).

[5] Klein (2023, p. 108-109) also correctly highlights the relation between this minority position and logical egoism (see V-Lo/Blomberg, AA 24:187; V-Lo/Blomberg, AA 24:151; V-Lo/Dohna, AA 24:740). Such logical egoism fosters authoritarian thinking, and 'thinking for oneself' cannot be equated with, nor provide any grounding for, such a stance. As Klein (2023, p. 108) observes: "although striving for enlightenment implies not accepting arguments from authority, it also entails a rejection of logical egoism, which can only lead to relativism and skepticism." Finally, being a rational agent, in its theoretical aspect, also means acting under the two other maxims of *Sensus Communis*, i.e., "to think from the standpoint of everyone else"; and "to think always consistently" (KU AA 05: 294).

[6] As Kant states (KU AA 05: 174; 176-179), these are completely different territories, which can only be unified by the faculty of judgment.

[7] Although it may seem relatively uncontroversial, this aspect is relevant for establishing the conditions of a Kantian theory of instrumental irrationality. As Martínez (2023) points out, Kant understood that not all representations are clear and evident; there are also obscure <*dunkel*> representations.

[8] In this definition, a rational agent may act against her own ends. Such an action would be considered irrational, yet she would still count as a rational agent. This means, of course, that a rational agent is not defined as one who never acts contrary to the ends she has set for herself, but rather as one who is capable of acting in that way. Within this not-so-intuitive framework, an "irrational agent" is not the opposite of a rational agent but instead a particular case of one. A person is called "irrational" when she performs a series of irrational actions, yet she is so only insofar as she also retains the capacity to act rationally. Accordingly, while the volitional aspect of rational agency consists in the possibility of acting either in pursuit of or in disregard of the ends one

With this initial account of what it means to be a (practical) rational agent, we can now introduce the concept of *irrationality* as having two possible meanings: *a failure* either in (i) the *intellectual aspect or* in (ii) the *volitional aspect*. As I intend to argue, *only volitional irrationality can be counted as a form of practical irrationality* and the intellectual aspect of rational agency will need a critical review.

2.2. Action

Distinct from other concepts, Kant offers few elucidative comments on how we can best present the concept of action. In the critical treatment offered in the KpV, there is no theory of non-moral or merely prudential actions. Since the investigation is specifically concerned with the foundations of morality, most of the commentary on the notion of action occurs in comparisons between good and bad actions, or, in other words, between actions that are legally and morally permissible and those that are not. At the beginning of Chapter 3 of the KpV (AA 05: 71-72), for example, he affirms that an action may contain legality, and among those, some may contain morality (if the will's determining ground is only the moral law itself). From this we can certainly conclude that (i) there are illegal actions (in the sense of forbidden according to reason), and (ii) there are legal actions that are not moral (actions performed in accordance with duty, but not from duty). Later in this chapter, he defends the thesis that some actions are called duties (AA 05: 80). These actions (duties) are conceived through the "exclusion of every determining ground of inclination." Again, by contrast, we can infer that actions which have some determining ground of inclination cannot be called duties[9].

For our purposes, it suffices to *concede hypothetically* (and also *provisionally)* that a given action can either fall within the moral domain or outside it (non-moral, as I will call it). We know well that within the moral domain, (i) actions may stem from a will whose maxim is universalizable to every rational agent, or (ii) they may stem from a will whose maxim, when universalized, entails a contradiction. The former, if also performed from duty (and not merely in accordance with duty), are moral actions. The latter are immoral and contrary to duty. The irrationality arising in this domain can easily be understood as a transgression of the moral law, and thus all type-(ii) actions are irrational. Hence, we understand here moral irrational actions as those that transgress the moral law. I believe this constructs a theory of irrationality without major controversial theses or conclusions.

Within the non-moral domain, there are actions whose subjective maxims cannot be taken as absolutely necessary and whose ends are not unconditional, but rather conditioned and relative to a particular agent. All of them would be counted as actions whose determining ground is inclination and, therefore, grounded not in a command of reason, but in a sensible element. This leads us to the question of whether a failure in an action of this field could be counted as a kind of irrationality; in other words, the central question is whether there is such a thing as irrational non-moral actions. As introduced previously, such a failure could hypothetically be (i) a failure arising from an intellectual error or mistake, or (ii) a failure caused by a volitional ground.

has previously set for oneself, any rational or irrational action is itself an expression of rational agency. When the action accords with the ends one has established, it is a rational action; when it does not, it is irrational.

[9] Another key feature of action is the existence of a double standpoint, which makes it possible to regard the very same action both as naturally determined and as transcendentally and practically free (KrV B XXVIII, B 566, B 579, B 585; KpV AA 05: 94, 99, 100, 104). I will return to this point in Section 4.2.3.

Once again, let us *assume provisionally* that prudential actions are a specific type of non-moral action, and they can be conceived as actions that promote (or at least attempt to promote) happiness. In other words, prudential actions are actions created by the faculty of desire taking the representation of happiness as its cause. In order to fully understand prudential actions, we must first understand the distinction between the two types of imperatives that Kant draws. In the next section, I will conduct an analysis based solely on the GMS to demonstrate that – even without the critical revisions Kant later introduced in the KpV and the KU – prudentially irrational actions cannot stem from intellectual errors.

3. The Theory of Imperatives in the Groundwork

3.1. Irrational actions outside imperatives?

Kant states that an imperative is the "formula of the command" which is, in turn, "the representation of an objective principle, insofar as it is necessitating for a will" (GMS AA 04: 413). Furthermore, "all imperatives are expressed through an ought" and they indicate that something ought to be done or omitted because it is good in a practical sense. Kant explains what is good in a practical sense and immediately contrasts it with the "agreeable" <*angenehm*>:

> Practical good <*Praktisch gut*>, however, is that which determines the will by means of representations of reason, hence not from subjective causes, but objectively, i.e., from grounds that are valid for every rational being as such. It is distinguished from the *agreeable* <*Angenehmen*>, as that which has influence on the will only by means of sensation from merely subjective causes, those which are valid only for the senses of this or that one, and not as a principle of reason, which is valid for everyone. (GMS AA 04: 413).

From this, it follows that every imperative operates according to the good, which is, in a practical sense, objectively good and stands in opposition to the subjectively good, namely, the agreeable. Therefore, this type of action falls outside the scope of imperatives, as it does not take the good as its foundation. Here arises the question: could irrationality be associated with what is subjectively agreeable but is not otherwise objectively good? If understood in a sense of *absence of reason*, then the answer is affirmative, since only what is practically good is good in a rational and objective sense. Thus, it is the case that guiding oneself by what is merely agreeable is irrational to the extent that there is no participation of reason in the course of action.

Certainly, this type of action (which aims at the agreeable) cannot be an action that conforms to any imperative. As Kant indicates in the footnote, it was proven in Section I of the *Groundwork* that an action done from duty cannot be one interested in the production of its object, but only for the sake of the action itself according to reason. It is also clear in the footnote that the agreeable is merely the anticipated object of the action. Thus, a will that acts out of interest in the object of the action is a will that acts according to a pathological interest (GMS AA 04: 413). Again, if we understand here the *absence of reason as irrationality*, then this type of action is certainly irrational. This certainly has a very different meaning from how we use 'irrationality' in *everyday language*, where it is generally understood more as *something contrary to reason* than as merely the absence of reason. Nonetheless, it does not seem to be the case that this type of action is irrational in the sense that we stated above, i. e., as a *volitional or an intellectual failure*. Textually, there is no indication that an action performed for the sake of producing the agreeable is irrational as contrary to a command of reason. One clear reason for this is the following: if an action is guided solely by the feeling of the agreeable and, therefore, bears no relation whatsoever to any duty, then it cannot be regarded as transgressing any form of command (or even coun-

sel) of reason. In such a case, reason is simply absent, and if it plays any role, it does so merely as an instrument of a will pathologically affected.

3.2. The Concept of Hypothetical Imperative

As we know, Kant distinguishes the imperatives between hypothetical and categorical. According to him,

> The former represent the practical necessity of a possible action as a means to attain something else which one wills (or which it is possible that one might will). The categorical imperative would be that one which represented an action as objectively necessary for itself, without any reference to another end. (GMS AA 04: 414).

I would like to stress the fact that both categorical and hypothetical imperatives are here stated as bounded to the duty and, therefore, to a normative account of action. However, as we shall see through Kant's critical revisions in the KpV and KU, this will no longer be the case. Furthermore, both imperatives are "formulas of the determination of action, which is necessary in accordance with the principle of a will which is good in some way" (GMS AA 04: 414). This passage makes it clear that a good will contains a principle guided by some source of imperative. As we know by the following line, this source of command could be either categorical or hypothetical. It is hypothetical only if "the action were good merely as a means to *something else*" while the categorical imperative represents an action that is good *in itself*. Kant then distinguishes hypothetical imperatives between imperatives of skill and imperatives of prudence (GMS AA 04: 414-415).

The imperative of skill is problematic because it guides action toward "any possible intention" (GMS AA 04: 415). It consists in finding the best means to achieve a given or conditioned end. The example of the doctor and the murderer makes this clear. According to Kant, the precepts for a doctor to cure someone and for a murderer to poison them with the same drug are the same and have the same value. Furthermore, the imperative of skill (or technical imperative) is possible according to an analytical principle of willing[10]:

> Whoever wills the end also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power. This proposition is, as regards the volition, analytic, for [...] the imperative extracts the concept of actions necessary to this end merely from the concept of the volition of this end [...] (GMS, AA 04: 417).

The imperative of prudence is assertoric, as it guides action toward a "real intention" (GMS AA 04: 415). According to Kant, beyond particular ends, all rational-sensitive beings have a common end, namely, happiness. Thus, the imperative of prudence is constituted by the "choice of means to one's own happiness" (GMS AA 04: 416). Furthermore, happiness is an (indirect) duty and an "ideal [of imagination]" from which "all inclinations are united in a sum" (GMS AA 04: 399). In terms of its necessity, the imperatives of prudence possess a greater necessity than the imperatives of skill, but not as great as the categorical imperatives, since only the latter have an unconditional necessity. As Kant indicates,

> The giving of counsel contains necessity, to be sure, but can be valid merely under a subjective, pleasing

[10] With this, Kant points out that we are not talking about the "execution of the action", but only about "how to think the necessitation of the will that the imperative expresses in the problem" (GMS AA 04: 417). As we will see, this is relevant because if one intends to develop a theory of Kantian irrationality based on the Groundwork, one cannot consider the necessitation that operates in the will but rather the execution of the action. This is counterintuitive and will be further elaborated upon later.

> <*gefälliger*>[11] condition, whether this or that human being counts this or that toward his happiness (GMS AA 04: 416).

In addition to carrying greater subjective necessity, the imperatives of prudence also entail a greater indeterminacy. The reason for this is that the imperatives of prudence are based on the concept of happiness, which (i) consists only of empirical elements and, at the same time, (ii) requires an "absolute whole, a maximum of welfare, (...) in my present and in every future condition" (GMS AA 04: 418). As is known from the KrV, the absolute of conditions cannot be known either transcendentally or empirically, and thus a sophism regarding the ultimate conditions of the world – whether with speculative interests or practical interests – is illegitimate and leads to a set of antinomic conflicts. In the GMS, this results in the impossibility of a finite being knowing for certain what one ought to do to be happy, since the totality of conditions is not given in the series of space and time. Happiness is not even an "ideal of reason," but merely one of the "imagination" <*einbildungskraft*>, and thus could not even constitute an object of practical reality (GMS AA 04: 418). For this reason, the imperatives of prudence are constituted only as "empirical counsels, e.g., diet, frugality, politeness, restraint," instead of "determinate principles" and they are in no way a "command of reason" (GMS AA 04: 418). Still in this same paragraph, Kant raises a serious doubt about the practical possibility of this imperative:

> the problem of determining, certainly and universally, what action will promote the happiness of a rational being, is fully insoluble, hence no imperative in regard to it is possible, which would command us, in the strict sense, to do what would make us happy (GMS AA 04: 418; emphasis added)

Later, however, he concedes that they would be "analytically practical propositions if one assumes that the means to happiness could be specified with certainty," and thus, "there is also no difficulty in regard to the possibility of such an imperative" (GMS AA 04: 419). Moreover, unlike the Categorical Imperatives, "their reality is given in experience" (GMS AA 04: 419-420).

3.3. Do Intellectual Failures Lead to Prudential Irrational Actions?

We now turn to the question of whether irrational actions can arise from intellectual mistakes inside the hypothetical imperatives. According to Merle (2023, p. 10-11), one source of instrumental irrationality lies in holding *incorrect beliefs*. Regardless of whether reason is present or absent, an agent may err in selecting the means to their ends. As Merle (2023, p. 10) notes, such mistakes can occur in two ways: (i) Misjudging the best means for given ends, or (ii) erring about whether one actually possesses the necessary means to achieve those ends.

The objection I wish to raise at this point concerns the possibility of instrumental irrationality – and, as I understand it here, the possibility of a prudential irrational action arising from an intellectual errors. I believe that the subject's error in these cases does not constitute prudential irrational action in either of the senses I proposed above. Let us take an example to illustrate this. Suppose I am walking down the street and suddenly feel thirsty. Therefore, my end is to quench my thirst. I consider buying water at the store or drinking the water I have at home. Since (i) my thirst is not great, (ii) I am close to home, and (iii) I want to save money, I decide

[11] It is important to state that the word "pleasing" <*gefälliger*> was changed by the editors in the Academy Edition to "contingent" <*zufälliger*>. Nevertheless, whether in the original or in the editors' version, Kant makes it clear that it is not the same subjective feeling that occurs prior to duty, namely, the "agreeable" <*angenehm*>. As noted, this term arises to distinguish an action according to an imperative from an action determined merely by the subjectively agreeable. On the other hand, "*gefälliger*" was used to differentiate the imperative of prudence from the categorical imperative, understanding the former as subjectively conditioned and the latter as unconditioned.

that the best course of action is to drink the water I have at home. However, when I arrive home, I discover that the city water company is conducting maintenance on my street and as a result I have no water at home. Since I am without water, I cannot drink it as planned. The conclusion is that the means I chose (namely, drinking water at home) proved to be erroneous because it was based on an incorrect *external belief*. Yet, it can be said that at the time of my decision, the action taken was based on the best possible conditions and, therefore, was rational. I was not actually in possession of the necessary means, but as shown, I had no way of knowing at the time that these means were unavailable to me. At that time, I took the best course of action based on the information I had and the critical analysis of the situation was sound. Therefore, not being in effective possession of the means does not imply the irrationality of the action, even if it implies an *external* wrong belief.

I call this an *external* wrong belief because I was led to a course of action over which I had no control or agency. If I am deceived by an external factor, can it really be said that I acted irrationally? It seems difficult to assert that a subject who has been deceived is irrational, because their course of action no longer depends on them. They acted based on the best possible presuppositions and conditions at the time, and it just so happened that, in this particular case, they were deceived by an external factor. I believe that this argument rules out the possibility of irrational action motivated by *external* factors or *wrong beliefs*.

Now let us examine the second possibility, namely, an *internal wrong belief*. Let us modify the thirst example a bit. In this example, let us suppose that I am already at home and have access to water. Let us also assume that I am thirsty, but in addition, I need to take a shower because I am sweaty from the walk. Finally, let us assume that I am short on time and therefore want to save time. In this case, instead of drinking filtered, good-quality tap water, I decide to drink water from the shower while bathing, because by doing this, (i) I quench my thirst, (ii) take a shower, and (iii) save time, as I am drinking water during the shower. A few days later, I end up getting sick, and the diagnosis indicates a likely causal link with the shower water I drank. Could it be said that I made an irrational decision? It does not seem so, because I fulfilled all the conditions and ends I set for myself at the time. However, something seems to suggest that the action was irrational.

The only condition under which the action described above could be considered irrational is to take into account not just this or that particular end within a given timeframe, but the totality of my ends across all time. This leads us to conclude that an *instrumental irrational action is impossible unless it considers the totality of ends across the entirety of time*. In the example above, my health is also a relevant end, and thus I did not use the best means to achieve the totality of my ends, given that this end, even if latent, was not fulfilled. This is a *necessary condition* for an action to be considered irrational with respect to the instrumental use of reason[12]. Furthermore,

[12] Since from this point on I carry out an analysis only from the point of view of practical rationality, and not theoretical rationality, it is nevertheless possible to consider here the case of an error stemming from theoretical irrationality. Let us suppose that I drank water from the shower because I saw a video of someone recommending this as a way to save time. Certainly, I did not make good use of my general faculty of knowledge, and I did not adopt a *critical attitude toward the information* presented to me. In other words, I used another's understanding to guide my life and remained in a state of minority. *From the point of view of theoretical rationality, I performed an irrational action*, since I did not carry out a proper procedure to investigate my beliefs. It is of utmost importance to emphasize that this differs from the notion of an intellectual failure in practical rationality, since the latter refers only to the failure in recognizing one's own self-imposed ends (for instance, a confusion). It is not about the procedure through which beliefs are assessed and subjected to criticism, but only about a misapprehension of the ends I desire (in this

as we have seen, imperatives of skill only concern a specific set of ends at a particular time, rather than the totality of ends across all time. An analysis based on the totality of ends would be subordinate to the imperatives of prudence and not to the imperatives of skill. Thus, from the standpoint of the execution of action[13] (where irrationality might *problematically reside*), there is no possibility of an irrational action arising from skill-based imperatives that are erroneously guided by false beliefs (whether internal or external). These imperatives lead us to a false positive irrational action, that is, an action that is in fact irrational if we take all the ends of an agent into consideration, but that seems to be rational if we consider only a specific set of those ends. Therefore, in a Kantian framework, this requires *conceptualizing non-moral irrationality through prudential imperatives rather than skill imperatives.*

In the case of imperative of prudence, it is difficult to conceive how someone could act contrary to such imperative. One clear statement that Kant does is that it is impossible to formulate universally valid objective rules as imperatives for happiness. This does not mean, however, that subjectively one cannot know through trial and error what brings him closer to or further away from his own happiness. Let us suppose I love studying and doing philosophy, but I am afraid of the future in terms of employment. Since I see that I am young, I might still give up and start over in another field, but I can also stay in the field and bet my chips on philosophy. Let us suppose I decide to stay and, in doing so, give up another career. This course of action will certainly lead to future outcomes, which are, at the moment, unpredictable. I may have good future outcomes in philosophy, or, even if I try hard, I might end up in a job I do not like and which, so to speak, makes me unhappy. Let us say this is the scenario and I end up in a job I do not like. Let's also say that this leaves me with a degree of happiness, on a scale from 0 to 10, of 5 points. Nevertheless, if I had taken a different course of action and decided to abandon philosophy and enter another profession, I still would not have the guarantee of a degree of happiness higher than 5. It could very well be the case that my degree of happiness is even lower, like 3 or 4. Thus, choosing one course of action or another cannot constitute an irrational action, as one cannot know what the best means to happiness (and its highest degree) are[14].

What can perhaps be said is that pursuing a course of action that I think may make me unhappy is irrational. Thus, if among the choices A, B, and C I reasonably believe that A and B have good chances of making me happy and C has no chance of making me happy, then it can be said that deliberately acting according to C is irrational. Therefore, an action is irrational in relation to the imperative of prudence only if we understand it not as a mistake to choose the best possible

case, conditioned ends). A possible objection might take the following form: in the case of the shower, could it not be said that I had a misunderstanding of the ends I desired? That is, I ignored health due to a miscomprehension of that specific end? In response, I consider that we must take into account, for the analysis, the totality of ends – something the skill imperative does not do, but only the prudential imperative can. The central point here is that, in the case of a rational agent's actions over time, one must always consider the totality of ends and not an isolated subset of specific ends. To isolate specific ends in such a way would amount to a grave error regarding the very notion of practical rational agency, as it would split the agent's ends arbitrarily – without reason – which is not only superfluous but certainly far from how we understand what it means to be an *unified agent* who desires or sets ends.

[13] Here, we are conducting an analysis based on the execution of an action derived from a technical imperative, and it is worth remembering that what constitutes each imperative, rather than the execution of each action, is simply its commanding relationship with the necessitation of the will. This is entirely consistent with the spirit of Kant's philosophy, since if we analyzed only the necessitation of each imperative in the will, we would end up merely reproducing the tripartite formulation of imperatives.

[14] Kant's examples are even better because they demonstrate that even desires for commonly valued goods – such as the pursuit of wealth, knowledge, longevity, and health – can lead to harmful ends (GMS AA 04: 418).

course of action, but merely as the deliberation in intentionally choosing a course of action that I conceive as having high chances of bringing me unhappiness. In this way, rather than irrational action according to this hypothetical imperative being a failure to maximize one's happiness, it is simply an act of choosing in favor of one's own unhappiness. This, however, is certainly not an intellectual error but a volitional failure. Such potentially irrational action would thus stem not from an intellectual mistake, but from a volitional one – that is, a case where despite clear representations, the agent's chosen course of action fails to maximize their own happiness. *Thus, we conclude that in no case do intellectual failures lead to prudential irrational actions*[15].

3.4. Setting conditions to Prudential Irrational Actions

As we can see, prudential irrational actions depend on two necessary conditions: (i) that happiness is conceptually distinct from instant pleasure[16], and (ii) that it is possible to rationally pursue happiness. If (i) is false, then no prudential action could be considered irrational, since the agent would always and immediately be choosing solely in favor of their own happiness (that is, every non-moral choice would simply be a choice aimed at increasing one's own happiness). If (ii) is false and reason does not play an active role in the pursuit of happiness – even if happiness and pleasure are indeed understood as non-identical – then it is not possible to judge an agent's action (or omission) in this pursuit as rational or irrational. Furthermore, I would like to stress that non-moral irrational actions should always be considered in light of all the ends of an agent and thus through prudential imperatives. As we have seen, taking into account only a specific subset of an agent's ends is problematic, insofar as it may lead to a *false positive* irrational action, i.e., an action that appears to be rational but is, in fact, irrational.

4. Two models of prudential irrational actions

If the argument presented in Section 3 is sound, we reach the following conclusion: *on a Kantian account, no practical irrationality arises from intellectual error*. We may now proceed to examine whether non-moral irrational actions can arise from volitional failures. For this investigation, I will employ two distinct interpretive models of Kant's theory of action. The first model, which I shall call the "negative prudential irrationality model", denies the possibility of irrational prudential actions. According to this model, all prudential actions must be considered at least as non-irrational. The second model – called the "positive prudential irrationality model" – makes the contrasting claim: prudential irrational actions are indeed possible. As we shall see, the for-

[15] This conclusion closely aligns with Klein's (2023) framework of irrationality. His approach rejects conceptualizing irrationality as mere failure or mistake, instead defining it as "a manner of thinking and acting grounded in erroneous principles" (Klein, 2023, p. 102).

[16] The example of a patient with gout serves as evidence for this distinction (GMS AA 04: 399). Kant tells the story of a gout sufferer who chooses to exchange his health for momentary pleasures, reasoning that he may not live much longer anyway. According to Kant, if happiness does not compel him through inclination, then it must compel him through duty. In such case, momentary pleasure must not prevail over the totality of one's inclinations – that is, over happiness itself. In this case, an irrational action understood as a transgression of the imperative of prudence is only possible insofar as it is an action that takes immediate pleasure as its foundation, rather than happiness. Perhaps this is what Kant meant when he referred to the "*agreeable*" as a purely subjective condition, in opposition to a *practical good*. As we shall see (4.1.4), for Timmerman (2022), both are happiness, and while the latter is the *broad* concept of happiness, the former is the *narrow* one. Moreover, we could say that happiness in its *broad* concept is a rational disposition toward the means of what could become a happy state of mind and therefore considers the means in the long term. This would represent the aggregate set of all satisfactions (an ideal of imagination). Meanwhile, in its *narrow* concept, happiness is the immediate satisfaction required by sensibility and thus considers the means in the short term. Here, the satisfaction arises from a single inclination.

mer model finds its basis primarily in Timmermann's interpretation of Kant, while the latter derives from Korsgaard's reading. As I shall argue, both accounts ultimately render non-moral irrational actions impossible – though they differ crucially in that Timmermann classifies prudential actions as non-moral, whereas Korsgaard treats them as steaming from unconditional reasoning and, therefore, from the CI.

4.1. The negative prudential irrationality model

In *Kant's Will at the Crossroads: An Essay on the Failings of Practical Rationality* (2022), Timmermann advances two central claims: (i) all prudential failures are *Socratic* – and thus intellectual failures – and (ii) all moral failures are *akratic* and consequently volitional. As established in Section 3, intellectual failures can never generate practical irrationality. Timmermann's central challenge, therefore, lies in demonstrating that prudential failures are exclusively intellectual, with no volitional component whatsoever. On his account, the HI is merely a principle that explains how non-moral actions come about in the world, not one that normatively evaluates those actions. His argument relies especially on Kant's reconsiderations in the two Introductions to the KU and in the KpV

In the KpV (AA 05: 25-26, 26n), principles of self-love are understood as principles that employ rules (or means) to achieve their ends. However, these principles are strictly theoretical. Kant's example is that if one wants to eat bread, one must build a mill. This is merely a rule of understanding that states the *intrinsic relation between cause and effect*. In the KU (AA 05: 171-173), Kant restates even more clearly the assimilation of rules of skill to theoretical principles and to theoretical philosophy. A common mistake, he claims, is to conflate the practical as a concept of nature (theoretical philosophy) with the practical as a concept of freedom (practical philosophy). The concept of "will" <*Willkür*> analytically contains the concept of the "practical", but we cannot know whether this "practical" refers to a rule of causality that derives from a concept of nature or from a concept of freedom. If it derives from the former, the principles are technically practical and belong to theoretical philosophy; if from the latter, they are morally practical and belong to practical philosophy.

What is of key interest to us follows directly from this distinction. Kant explicitly states that technically practical rules belong to theoretical philosophy, and that all rules of skill and all rules of prudence are technically practical. In other words, Kant is asserting that hypothetical imperatives do not belong to practical philosophy because their rule of causation does not originate in freedom, but in nature[17].

This point may seem subtle, but its philosophical implications for Kant's system of action are profound. When the concept guiding the will in the hypothetical sphere is nature, it follows that no free agency is possible. In fact, the entire context of the first part of the KU's Introduction is meant to distinguish theoretical from practical philosophy. While theoretical philosophy

[17] In the First Introduction (AA 20:199-200), Kant's philosophical revision regarding HIs becomes even more explicit. He states that it is time to "correct a mistake I made in the Groundwork of the Metaphysics of Morals." In his reformulation, Kant considers that he should have called the imperatives of skill technical imperatives and included the pragmatic (or prudential) imperatives within the set of technical imperatives. They still retain a proper name because, whereas technical imperatives (insofar as they are imperatives of skill) have a purpose <*Zweck*> that is given or presupposed as known, the pragmatic imperatives must still determine what this purpose consists in (namely, happiness) and not merely deliberate on how to achieve it. In other words, as an end, happiness is given to the agent, but its determination remains uncertain.

is grounded in the concept of nature, practical philosophy – through the concept of freedom – enables the possibility of a metaphysics of morals. Such a philosophical enterprise is only possible because free action is possible for the human will. When Kant isolates hypothetical imperatives to the domain of theoretical philosophy, he is implicitly denying that any free action can originate from them. Now, I would like to stress two major implications that follow from this critically revised version of the HI presented in KpV and in KU.

4.1.1. Determinism of the action

The first and most obvious implication is that one could not act differently when acting on the basis of HIs alone. Since HIs are conceived as concepts of nature, there is no space for freedom to be imperative or authoritative in the actions that arise from them. To be more precise, practical reason always plays a role in the process of setting ends. However, reason can never override the power of sensibility (and therefore inclination) in this realm. The key concept attached to the practical meaning of will is not freedom, but nature and its natural mechanism. This might not be convincing enough, so I would like to offer another argument in support of this claim. Instead of appealing to the critically revised version of the HI that Kant presents especially in the KU, let me defend this view only by reconstructing Kant's position in his theory of action as laid out in the GMS and KpV.

Kant states in the Third Section of the GMS (AA 04: 447) that "a free will and a will under moral laws are the same. Thus if freedom of the will is presupposed, then morality follows together with its principle from mere analysis of its concept." Let us first understand why morality always entails freedom. When my dog attacks my cat, I can certainly scold him for it. I might even say to him: "Bad dog! You're grounded for misbehaving." However, deep down I know he is not morally responsible for his behavior. I discipline him in a specific way to reinforce the behaviors I want from him (like not attacking my cat, for example). The situation would be completely different if, instead of my dog, a friend of mine attacked my cat for no reason. At the very least, I would certainly reconsider our friendship or even call the police.

What is the key difference at work here? It is that animals are not free and therefore their actions cannot be morally evaluated, whereas humans are. This distinction is grounded (at least in modern times) in our common-sense morality. As we can see, acting morally clearly entails acting freely. That is why Kant says that the *ratio essendi* of morality is freedom (KpV AA 05:04). Without freedom as the metaphysical foundation, morality would be nothing more than a "chimerical idea" or a "figment of the mind" (GMS AA 04: 445).

What is more difficult to grasp is why every free act is also a *morally relevant* or a *moral-type* act[18]. The answer lies in the relation between sensibility and reason. In its negative meaning, freedom can only be conceived as a detachment from sensibility (GMS AA 04: 446). One who acts solely according to sensibility acts exactly like an animal. No morality is derived from such action, for it is conceived merely as a mechanically caused response to the interplay of sensible forces. One detaches from sensibility only insofar as one considers one's maxim not as determined by sensible inclination, but as something that could be universally willed.

[18] Here I take morally relevant and from a moral-type as the same. As Hill (2013, p. 20) points out, immoral actions, i.e., actions against the unconditional duty are also free (otherwise they would not be imputable). Nevertheless, the *expression of freedom* can only be conceived as actions performed in accordance with the moral law.

We cannot know freedom apart from the moral situation. It is only through morality that we initially become aware of freedom. This is why Kant calls the moral law the *ratio cognoscendi* of freedom (KpV AA 05:04). Outside the moral sphere, we know nothing about freedom, for that would amount to trying to discern an idea beyond the bounds of practical experience. Detached from its practical use, the attempt to know freedom theoretically is the product of an undisciplined reason that falls into transcendental illusion. Without morality, the practical domain would be understood merely as a theory *explaining* actions. As Korsgaard (2008) points out (and as we will see in more detail later), it would not be possible to assess an agent's actions, and thus we would not be able to say whether a given action was right or wrong. In other words, normativity arises only insofar as there is a higher-order evaluative principle that allows actions to be judged as correct or incorrect. Through HIs alone, we would never reach the concept of freedom, for there would be no normativity enabling us to judge actions as coherent with any external principle. It is only the moral situation that grants legitimate entry for the concept of freedom into the realm of practical experience.

But does this prove that all freedom is exclusively moral and can never be instrumental? One might suggest that we gain initial access to freedom through morality but that freedom is not thereby limited to it. In response, I would ask: how could freedom pertain to actions that are not moral? For what reason should we believe such a claim? It seems like a dogmatic assertion that attempts to extend the use of a concept into illegitimate terrain. In other words, it may be metaphysically possible that the spontaneity of practical reason extends beyond morality, so that we might also be free in the purely instrumental domain (i.e., apart from morality). However, this would be solely a problematic proposition from the standpoint of theoretical philosophy. In the Kantian spirit, we must ask what would lead us to treat such a proposition as possessing practical validity – that is, as being an assertoric judgment for the practical domain. Such a move would be an illegitimate leap, for we have no grounds to claim that we know the idea of freedom outside its moral use. Hence, this metaphysically problematic judgment cannot become an epistemically assertoric one within the practical domain.

Now we can clearly see that (i) if an action is conceptually conceived as free, it is always also conceptually conceived as morally relevant. Thus, (ii) every free action is a morally relevant action. (iii) If an action is not free, then there is also no free agency. And (iv) where there is no free agency, there can be no irrationality. (v) If we represent an action as purely instrumental, then it is certainly not moral. It follows that (vi) *if an action is purely instrumental, it cannot be considered irrational, since it is not morally relevant and therefore not free*[19].

4.1.2. Analyticity vs. syntheticity

Another significant shift introduced by Kant's critical revision of the HI, particularly in the KpV and the KU, concerns how we can understand the concept of analyticity in relation to HIs. As Timmermann (2022, §19, p. 56) states, HIs do not require any motivational status to determine the agent's will. It is precisely because the agent is already determined by a specific end that she finds herself motivated to will the means to achieve it. Since there is no end beyond sensibility, the agent desires the end immediately and without any volitional resistance. As Timmermann

[19] Here of course I am considering one hidden premise, i.e., that every action can only be thought as purely instrumentally if and only if it arises directly from HIs. The CI cannot generate purely instrumental actions.

(2022, p. 59-61) explains, the hypothetical imperative, being an analytic practical principle, requires only that the agent has a sufficient theoretical understanding of the connection between means and ends in order to act. Once the agent knows all the necessary means to attain her end, she acts in accordance with the desired end. There is no room for alternative courses of action or genuine choice. To say that the proposition is analytic means that the agent acts immediately upon the representation of her end, as long as the means to it are sufficiently known[20].

Another way to understand the distinction between synthetic and analytic imperatives is by comparing them with the divine will. A divine will acts according to the moral law in a purely analytic manner – that is, it is sufficiently and always motivated merely by the representation of the law itself. In Kant's terms, the objective necessity of the moral law is identical to the subjective necessity of this law. In contrast, human beings have a will that is not purely intelligible but also sensibly affected. For this reason, the moral law is not immediately taken as the foundation of our actions. Thus, for us, the moral law can only be represented through a *synthetic* proposition, not an *analytic* one. The *syntheticity* of the CI indicates that sensibility stands between us and the representation of the moral law and, for this reason, the moral law is not immediately sufficient to motivate our actions. We are always divided into a two-fold possibility, whereas for a divine will there is always only one possible mode of acting. As previously argued, because we possess freedom of the will, this two-fold possibility is not merely illusory, but real. If we represent ourselves as lacking freedom of the will, two options remain regarding the status of the moral law:

(i) Action would always follow analytically and immediately from the representation of the law (as in the case of divine wills); or

(ii) The moral law would be empty for the will – even if it were represented, it would not be possible to act against sensibility. In this second scenario, action would be entirely determined by the sensible nature of the agent.

The first case describes divine wills; the second corresponds to the empiricist thesis that Kant criticizes from the KrV onward, especially in the third section of the GMS. In contrast, the synthetic character of the CI signals the obstacle that exists between the representation of the moral law and the sufficient motivation for producing the object demanded by it (which, according

[20] Timmerman points out that the normativity of HIs lies in the requirement of knowing a series of theoretical rules in order to achieve ends. Ignorance of the means implies failure in the fulfillment of ends. The notion of an imperative as a rule that can possibly be violated by the agent is thus preserved, since the agent may fail to follow the theoretical rules that he does not know in order to reach the desired end. Such a violation, however, does not necessarily occur knowingly or willingly. As becomes clear, this error is merely intellectual rather than volitional, for it does not indicate any lack (or weakness) of motivation of the will, but only a will that does not know the necessary means to achieve its ends. If I want to bake a cake but do not know the ingredients, then there is a transgression of the hypothetical imperative insofar as I want the end but do not know the means. Nevertheless, even within the commentator's argumentative framework, this seems mistaken: if I want the end but do not know the means, I cannot truly want the end, for I cannot want the means. If I want to travel to Switzerland but have no money, then I merely have a wish <*Wunsch*> and not a desire <*Begierde* (*appetitio*)> (Anth AA 07: 251). The latter is connected to the capacity (or *faculty*) to desire <*Begehrungsvermögen*> (MS AA 06: 211), whereas the former, when it does not become a ground for the faculty of desire, is a peevish wish <*launische Wunsch*> (Anth AA 07: 251). The HI entails the following structure: If I want to go to Switzerland, then I also want the money to do so; and if I am not willing to pursue the money I lack, then I do not truly want to go to Switzerland. I might wish for it, but I do not genuinely desire it. Wishes arise in the mind and may by themselves motivate action. Desire, however, is expressed in the world as a form of causality. I may have many wishes, but I desire only one thing (and therefore must abandon contradictory wishes). As I understand it, this allows only a *post factum* explanation of what the agent really desired in the past. This is very close to how Schopenhauer (1960) conceptualized the will. In his account, I can only know what I truly wanted in the past and never know exactly what I want in the present. Willing can only be seen in the world retrospectively, through actions that have already taken place.

to KpV, is the good). The agent can therefore either succeed in following the moral law or fail morally. This dual possibility, opened by the freedom of the will, is crucial for moral imputation. Without it, morality would be no more than a "figment of the mind" or a "chimerical idea" (GMS AA 04: 445). As we can see, synthetic propositions are those in which there is a volitional obstacle between the representation of the end and the action that would make that end real.

Furthermore, the analytic character of HI also correlates with the impossibility of contradiction as presented by Kant in the KrV. If in analytic propositions the predicate and the subject are thought through a relation of identity, then to affirm one while denying the other generates a contradiction. It would be like saying, "It is not true that every bachelor is unmarried." This statement is contradictory because the concept of "unmarried" is already contained within the concept of "bachelor." Someone who genuinely tries to defend the truth of such a proposition surely does not grasp at least one of the two concepts involved.

In the practical field, this would be like wanting to make coffee but refusing to fetch the coffee grounds, filter, and dripper from the drawer. This is a practical contradiction and, in this case, if one does not fetch the filter, the grounds, and the dripper, then he does not really want to make coffee. The act of making coffee and fetching its required elements from the drawer is a single act, and thus there are no extrinsic or instrumental desires derived from the desired end (making coffee). The means are immediately desired from this end and, in this way, the agent intrinsically wants to get the grounds, the filter, and the dripper. If he receives some friends at home and says that he will make coffee for them, but instead of using the coffee grounds he grabs the guarana powder that was right next to it in the drawer, we would surely interpret this oversight as a mistake or failure in the faculty of judging the appropriate means (that is, as an *intellectual* error)[21]. I find it difficult – following Timmerman (2022 §21–23) – to think we would judge the case as one involving weakness of will. In other words, we would judge the case as involving an intellectual mistake, but it would not follow from that that one acted irrationally in taking the means to his end. Thus, the failure to comply with the HI means solely an intellectual error on the part of the agent in understanding the actual causal relation between means and ends. If at time T2 I realize that means M2 were worse than M1 for achieving end E1, then there was an intellectual error, which, however, does not imply an irrational action. The HI becomes a theoretical apparatus for evaluating actions from a non-moral and, in a certain sense, non-volitional standpoint[22].

Analytic practical propositions indicate the immediate necessity between an agent's ends and the action undertaken to achieve them. Given their analytic relation to the moral law, the divine will can never fail to conform to it, and thus no command is required. The same holds when we, as human beings, regard ourselves as situated outside the moral domain. Kant understands HIs as analytic propositions, since anyone who genuinely wills the end faces no volitional obstacle to also willing the means. This, however, does not exclude the intellectual obstacles we may encounter in relation to the means required for our ends.

This is especially the case with happiness. We cannot be certain whether M1, M2, … Mx will

[21] This is very similar to the conclusion we reach in section 3.3, i.e., irrational actions can never stem from imperatives of skill.

[22] In Timmerman's words (2022, p. 57): "We violate hypothetical imperatives if and only if we fail to grasp the instrumental connection they contain, i.e. out of ignorance. If we act differently when we are clearly aware of the means and it is known to be at our disposal, we are left to conclude that we do not desire the end with the requisite force, that we want something else more than the end we think or say we prioritize; and this hardly counts as rational failure, even less as practical irrationality."

make us happier, and we may become intellectually confused in this realm. We choose M1 at T1, later realize it was not as fulfilling as we expected, and then switch to M2 at T2. At T2, however, we might reflect again and conclude that the former was actually better, leading us to return to M1. Let me illustrate this: suppose someone pursues a career as a teacher, abandons it to make a living as an artist, and then realizes that being a teacher was in fact more satisfying, returning to the classroom. This shifting of means to achieve happiness is constitutive of human life and, according to this model, the change is due to intellectual errors made along the way rather than to volitional weakness. This is why Kant states in GMS (AA 04: 418) that imperatives of prudence are, at least with regard to the determination of their object, impossible.

4.1.3. Abandoning ends

Yet, making an intellectual mistake cannot explain how we abandon ends in our daily lives. The coffee-guarana example is illustrative in showing how, given a specific end, an error in the choice of means does not imply irrationality. It does not explain, however, the abandonment of ends I once set for myself in the past. Instead of the explanation that there may be means-end irrational actions, Timmerman defends the hypothesis that the abandonment of an end is merely the abandonment of a means in relation to the totality of ends, that is, happiness. In other words, abandoning a specific end simply means abandoning an end that no longer makes sense within the overall calculation of happiness. Put differently, this end no longer promotes the same degree of happiness as was once expected. In this case, according to Timmerman, there is nothing irrational about the abandonment of ends, and a theory of irrationality could not be derived from such cases.

Let us take an example to illustrate this. Suppose Gabriel is a young university student who has set himself the goal of improving his academic performance this semester and, therefore, plans to engage more deeply with assignments and activities that contribute points in the courses he is enrolled in. Moreover, he has also decided to take better care of his health this semester and is thus aiming to lose fat and gain lean muscle through a specific diet. Now, at the end of the semester, he knows he must focus especially on writing an essay for his ethics class and maintaining a healthy diet. On Thursday morning, he thinks: "This week I haven't done everything I needed to. But today will be different. Today I'll commit." Around 11:30 a.m., contrary to what he had planned for himself, he no longer wants to have lunch and then write his paper; instead, he now prefers to play video games. *He is fully aware of the consequences*: (i) he will feel hungry and weak throughout the afternoon; (ii) he will not meet his nutritional goals for gaining lean muscle; (iii) he will have to write the essay later in the week; and (iv) the quality of the essay may suffer, possibly lowering his grade. Timmerman's (2022, p. 58) account would say that Gabriel simply abandoned the end of writing his essay because he now considers that skipping lunch and playing video games all afternoon would yield a greater total amount of happiness. Gabriel would be committing a HI mistake only and only if he considered something false as if it were true (for instance, that not having lunch would not make him feel hungry).

For Timmerman, the intellectualist position readily explains why Gabriel acted as he did: Gabriel now considers that his overall happiness is greater than it would have been had he followed through with his previous plans. In this way, Timmerman (2022, p. 58-59) reduces the HI to an explanatory principle of non-moral actions, one that does not serve to normatively assess

the rationality of actions, but merely to indicate and clarify *post factum* what the agent's actual reasoning was at that moment in terms of maximizing his overall happiness. The advice, for example, "You should not make a lying promise, so that if it were revealed then you would lose your credit" (GMS AA 04: 419), contains a conditioned duty, but one that can ultimately be abandoned by the agent, even without knowing for sure whether such abandonment will bring about greater happiness or unhappiness. Timmerman (2022, p. 58) claims that it will never be possible to know whether the abandonment of one particular end in favor of another has, in fact, generated more or less happiness than would have been the case in the counterfactual scenario.

4.1.4. The indirect duty to happiness

Although well-articulated and textually well-grounded, this negative model faces a very specific challenge. In a few passages of his work, Kant refers to a duty to promote one's own happiness (GMS AA 04: 399; KpV AA 05: 53; MS AA 06: 387-388). This is potentially problematic for the negative model, insofar as a moral duty to act in ways that enhance one's own happiness can only make sense if actions concerning happiness are understood as involving a rational aspect. Thus, happiness cannot, in this case, be a concept that immediately drives the agent toward the most pleasurable action (within the totality of the sensible play), but must instead be understood as a possible object that the agent may choose not to pursue in favor of another. In other words, if the agent must choose happiness as a duty, what would be the alternative choice? What other course of action remains when one fails to fulfill the duty to act for one's own happiness? As we have seen, one possible response would be to dissociate the feeling of the *agreeable* from the *practical good*, and to understand happiness as residing only in the latter. However, another condition is necessary for duty to arise. A duty can only emerge if the agent has the possibility both to follow and to transgress it. Without this dual possibility, there is no choice, and therefore no duty. This implies that the agent could deliberately choose against their own happiness – something that, up to this point, the negative model cannot accommodate.

According to Timmerman (2022 §25, §36), imperatives of skill are sometimes also necessary for the use of the CI. That the happiness of others is a duty is clear from the CI. However, the complexity of the world makes its application difficult. HIs enter here as tools for the application of the duty of beneficence. The case of the indirect duty to promote one's own happiness also falls within this scope. Promoting one's own happiness is merely a means to attain the moral end, and thus the hypothetical imperative to promote one's own happiness is only a tool used by the agent in order to reach that moral end. *Instrumental reason is here employed as a tool of moral reason and, outside of it, possesses no normativity*[23]. As we will see, Korsgaard will defend the same conclusion.

In this way, perfect duties have a clear application in the world. Imperfect duties, however, may require hypothetical imperatives to guide action. Normativity stems solely from pure practical reason, and instrumental reason is subordinated to it in this specific case. Mistakes in the execu-

[23] In contrast to Kohl (2018), Timmerman argues that not all hypothetical imperatives derive solely from one's own happiness. However, Timmerman maintains that the same instrumental procedure operates both (i) in the pursuit of one's own happiness (which is not in itself a duty) and (ii) in the promotion of others' happiness (which is a duty). For instance, taking a taxi to go to the bank and pay my debt requires the use of technical imperatives for the fulfillment of this duty. The action will have moral worth if it is carried out by taking reason as the sufficient cause of the action – that is, by acting from duty – and it will lack moral worth if the action is merely in accordance with duty.

tion of imperfect duties must be understood as errors in instrumental deliberation, rather than volitional moral failures.

The case of the gout sufferer (GMS AA 04: 399) pushes the difficulty raised here to its limit. The gout sufferer faces the following problem: do I pursue the ideal of happiness (through health) and thereby renounce immediate pleasure, or do I pursue immediate pleasure and thereby renounce the ideal of happiness? The ideal of happiness, let us recall, is merely an ideal of the imagination, and the "sum of all inclinations" can never in fact be realized in this world. This is the *broad* sense of happiness. In contrast, the *narrow* sense of happiness is the pursuit of a single inclination at the expense of the imaginary sum of all inclinations. The dilemma lies in the fact that the agent does not truly know what will make him happier: perhaps it is the *immediate* (*agreeable*) pleasure, or perhaps it is the *ideal* of happiness (*practical good*). He acts only according to the guiding principle of pleasure and, thus, will take the course of action that he sees as having the greatest chance of making him happy. At this point, there is no normativity yet, and he acts merely in accordance with *felicific happiness*. So far, neither option is rationally preferable or dismissible.

Following Timmerman (2022, §36), we can argue that normativity arises only insofar as the agent encounters the indirect duty to promote material conditions that facilitate the fulfillment of duty. Among these conditions there is a subcase of happiness, namely, health. In this specific case, happiness is taken as a means for promoting the unconditioned duty. That is, *when I am confronted with duty*, I can then choose between a *broad* or *narrow* conception of happiness. Outside the representation of an unconditioned duty, such a choice cannot be made deliberatively. Agents may always be playing the game of choosing between the ideal of happiness and happiness as mere pleasure, but this game is not grounded in reason. Inclination has the final word. Under the conditions established in section 3.4, criterion (i) is satisfied whereas criterion (ii) is not. However, when reason enters the game with the concept of a duty (in specific cases such as not placing oneself in a position that undermines the fulfillment of morality), the agent then has a real choice between the imaginative ideal of happiness and momentary pleasure. This choice between happiness and momentary pleasure does not occur in every case, but only hypothetically in cases where happiness is taken as a means to the fulfillment of some unconditioned duty. This is of great importance, as it explains why Kant affirms that happiness *may* become a duty in some cases, yet he never claims that it *is* a duty. Thus, *the standard position is that happiness is not a duty, but in particular and specific cases it may become an indirect duty*. The gout sufferer's mistake in choosing momentary pleasure instead of happiness as an ideal becomes a volitional error only insofar as the concept of duty comes into play. Without it, the choice between general satisfaction over time and momentary pleasures is simply a choice guided analytically by what the interplay of faculties presents to the agent as producing the greatest happiness. In the case of the gout sufferer, were it not for the representation of the duty to maintain a position conducive to fulfilling duty, he would not have acted irrationally.

4.1.5. Synthesis of the negative model

As we shall see, this model has the historical and interpretative advantage of accounting for all parts of Kant's work relevant to the issue. It can adequately explain the critical revisions undertaken by Kant in KpV and KU without abandoning major theses stated in GMS, such as the

radical distinction between the analyticity of the HI and the syntheticity of the CI. Philosophically, its account of non-moral actions closely resembles the way Hume presented his theory of instrumental rationality. This is entirely accurate, and yet it does not reduce Kant's theory of action to a kind of Humeanism. This model merely presents his theory of prudential actions as ultimately determined by the sensible aspect of the agent. I recall here that this by no means implies that reason is absent, but only that it does not override inclination in this domain. It is also not true that instrumental rationality is absent; rather, such rationality exists only insofar as it is subordinated to moral rationality. Moreover, when guided merely by HIs, the errors agents may commit are by no means expressions of irrational actions. They are merely intellectual errors that can be analyzed *post factum*. As long as it is subordinated merely to a technical imperative, happiness is pursued mechanically. *Therefore, non-moral actions (including those that concern happiness) are not susceptible to volitional errors, which are the only ones that truly give rise to irrationality. Although historically accurate, this model does not allow us to construct a Kantian theory of instrumental irrationality.*

4.2. The positive prudential irrationality model

Whereas the negative model interprets Kant's theory of prudential action as a species of Humeanism, the positive model asserts the very opposite: instrumental reason – including prudential actions – is normative because it is grounded in unconditional reasoning. In this positive model, it is practical reason, rather than inclination, that ultimately governs prudential action. Korsgaard is perhaps the foremost defender of this positive model. In her essay *The Normativity of Instrumental Reason*, she argues that (i) instrumental reason alone cannot fully account for how we deliberate about our ends and, therefore, (ii) it is insufficient on its own to ground a theory of instrumental irrationality. To address this, (iii) unconditional reason must play a significant role in the process of setting and acting upon ends.

4.2.1. Establishing the argument

Korsgaard's argument begins with the thesis that both instrumental and moral reason share a common source: "the autonomy or self-government of the rational agent" (Korsgaard, 2008, p. 23). To defend this view – just as Kant did – she must criticize two major philosophical traditions: empiricism and rationalism. To illustrate her criticism of empiricism, which she takes Hume to exemplify most clearly, let us turn to Howard's case (Korsgaard, 2008, p. 39). Suppose Howard, a man in his thirties, wants to live past the age of fifty and must take a series of injections in order to do so. However, Howard has a deep aversion to injections and therefore refuses the treatment. Let us further assume that Howard makes no intellectual error: he fully understands the consequences of his omission. According to Hume's account – as well as Timmerman's version of Kant – we cannot say that Howard is acting irrationally: he is simply acting on what he most strongly desires. His omission is merely a means to an end that he wills, namely, avoiding injections. As we have seen – and as Korsgaard (2008, p. 40) rightly notes – under the negative model, he cannot even be said to violate the principle of instrumental reason. As Korsgaard summarizes:

> The problem is coming from the fact that Hume identifies a person's *end* as what he *wants most*, and the criterion of what the person wants *most* appears to be what he actually *does*. The person's ends are taken to be revealed in his conduct (KORSGAARD, 2008, p. 40).

In this distinctly Schopenhauerian fashion, this account of prudential action can never be evaluated in normative terms. By itself, the prudential (instrumental) principle of adopting the best means cannot generate normativity. Normativity arises only insofar as we take the agent to be capable of freely committing to or abandoning ends. Rationality – and thus irrationality – emerges only when the instrumental principle is supported and grounded in a principle for choosing ends. That is why "Hume has no resources for distinguishing the activity of the person herself from the operation of beliefs, desires, and other forces in her" (Korsgaard, 2008, p. 45). These sensible elements may be sufficient, under the negative model, to explain human action entirely – but certainly not to evaluate it normatively.

In its revised form, the instrumental principle has, according to Korsgaard (2008, p. 46), the following structure: "if we will an end, then we ought to will the means to that end"[24]. Here, an important comment must be made. Korsgaard claims that the distinction between will (or desire) and wish is not available within the Humean empiricist framework. While this may well be the case for Hume, it is certainly not the case for the way the negative model was previously presented. As exemplified by the case of Switzerland, we can distinguish a mere wish from a true desire – but only through *post-factum* analysis. If Korsgaard is right, however, the positive model enables a distinction between desire and wish *ex ante*, i.e., prior to the action's actual occurrence. If volition is attributed to the capacity to freely adopt ends (even prudential ones), then we are no longer dealing with a will that merely results from the conflicting interplay of multiple wishes. This means that we can, at the very least, regard the agent's action as the product of a free will – and, therefore, that the agent is internally capable of distinguishing between what she merely wishes and what she truly wills. Within her own mental state, the agent can identify *ex ante* what she genuinely wills and what she merely wishes.

In this formulation, the 'ought' necessarily presupposes the possibility of transgressing this very ought. Beyond being a rational demand, it is a rational demand that can be freely violated – that is, willingly left unfulfilled by the agent. Korsgaard (2008, p. 51) provides one way to represent this structure syllogistically

> Whoever wills the end wills the means insofar as he is rational.
> I will the end.
> Therefore I will the means insofar as I am rational.
> Therefore I *ought* to will the means.

However, as she rightly observes, it is impossible to provide a non-trivial explanation for why an agent understands as necessary the adoption of means to their ends, since rationality (even in its volitional sense) is already presupposed in the premise. Thus, we cannot properly grasp the agent's *motivational foundation* for committing to the means for his end. This constitutes a

[24] This formulation of the instrumental principle is textually grounded especially in the following passage: "How an imperative of skill is possible requires no special discussion. Whoever wills the end, also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power." (GMS AA 04: 417). This passage has generated numerous readings of the hypothetical imperative as a normative or prescriptive principle, rather than a descriptive one. Timmerman's (2022, p. 62) interpretation of this passage, who presents the HI as descriptive, is offered in the following terms: "It remains true that imperatives apply only to those rational agents that can violate them; but it does not follow that they must be able to violate them knowingly or willingly. The parenthetical remark is perfectly compatible with there being a multiplicity of technical rules that we find hard to follow because we have trouble finding out what they say." A less structural and more historical interpretation would simply state that Kant revised this in the KpV and in the KU. Surprisingly, Korsgaard (2008, p. 48) calls the parenthetical part a caveat and notes that this would raise further problems for Kant's argument.

pivotal moment in her argument, as Korsgaard must partially abandon a distinction central to Kant – namely, the analytic-synthetic distinction.

4.2.2. Exegetical and historical limits of the argument

This brings us directly to the exegetical and historical limits of Korsgaard's argument, when Kant's own work is taken as the standard. As previously stated, Korsgaard must partially abandon Kant's analytic-synthetic distinction in order to reinterpret the hypothetical imperative. According to her, there is a "historical explanation for what has gone wrong here" (Korsgaard, 2008, p. 51). She argues that Kant falls into a kind of realist rationalism (*à la* Leibniz and Clarke) when he considers that the moral law would be analytic for us if our will were perfect. If the moral law is expressed through a series of truths or facts external to our agency, then the internalist criterion for why we ought to act rationally can never be fulfilled. And if that criterion is not satisfied, no coherent theory of rationality can be sustained[25]. In her new and final formulation of the instrumental principle, Korsgaard presents it as the commitment an agent makes to the end they set for themselves. As Korsgaard (2008, p. 58) explains:

> [F]or the instrumental principle to provide you with a reason, you must think that the fact that you will an end is a reason for the end. It's not exactly that there has to be a further reason; it's just that you must take the act of your own will to be normative for you.

This formulation, however, differs very little from the third formula of the CI provided by Kant, as Korsgaard (2008, p. 58) herself acknowledges. The analyticity of the principle is certainly called into question here[26]. What gradually takes place in her argument is the incorporation of the HI into the CI. As a result, the HI – and its normativity – becomes a special case of the CI. This is certainly a philosophically rich and fruitful construction. As Korsgaard points out (2008, p. 60 fn), one of the conclusions of her essay is that there can be no instrumental norms unless there are also unconditional ones.

However, this interpretation raises historical and exegetical concerns. The reduction of all practical rationality to the CI is not, in itself, a serious problem. The greater challenge lies in demonstrating why we cannot simply suppose that Kant did not regard the instrumental principle as rationally guided – that is, subject to normative assessment. In other words, why should the internalist requirement regarding the instrumental principle concern us? To support her case, Korsgaard (2008, p. 60-61, 67) suggests that there is a contradiction in those who fail to act on the instrumental principle: if you do not act on this principle, you cannot claim to have a unified will. Yet, in Kantian terms, there is no clear reason to suppose that such unity must be grounded in the instrumental principle. One could argue, instead, that the unity of the will derives solely from pure practical reason – that is, from acting according to the CI alone. This could be a potential response by those defending the negative model.

[25] This concern with the internalist criterion already appears in *The Sources of Normativity* (Korsgaard, 1996, p. 14; 46).

[26] She attempts, to some extent, to preserve the analyticity of the principle by asserting that "to will an end just is to cause or realize the end, hence to will to take the means to the end," and thus that "the instrumental principle is constitutive of an act of the will. If you do not follow it, you are not willing the end at all." This, however, can only be seen as a metaphor for how Kant described the analyticity of practical principles. In Korsgaard's position, the analyticity of the instrumental principle derives from a double exclusionary condition: either you will the end, or you abandon the instrumental principle. Yet I could not rationally abandon the principle, since it is constitutive of the will. For Kant, however, this formulation would count as a synthetic one. If I can – albeit at the cost of being irrational – freely abandon the principle, then my relation to it is synthetic. As Kant maintains, the analyticity of a principle derives from its irresistible motivational force upon the will. An analytic principle, by definition, can never be transgressed.

Furthermore, Korsgaard states that "Kant's account" of "the principle of prudence as a hypothetical imperative (...) is in need of revision" (2008, p. 29 fn). She also claims that it is necessary to offer a kind of "deduction" for HIs, analogous to the one Kant provides for the CI (Korsgaard, 2008, p. 29). The problem is that Kant explicitly states that all hypothetical imperatives are grounded in experience and do not require any special deduction (GMS, AA 04: 419-420). At a certain point – and this is of major significance for a historical interpretation – Korsgaard appears to abandon the analytic-synthetic distinction altogether. She concludes: "I am inclined to think that my argument shows the distinction to be less important than Kant thought" (2008, p. 62-63 fn).

4.2.3. Expanding the model

Building upon Korsgaard's considerations – yet also going beyond them – I would now like to show how the positive model could be expanded. Following Korsgaard (2008, p. 67-68), the first premise we must accept in order to develop this expanded model is that the hypothetical imperative is merely "an aspect of the categorical imperative," and thus that "there is only one principle of practical reason, and it is the categorical imperative."

The advantage of this model is that it can certainly provide a better account of Kant's considerations regarding the indirect duty to promote one's own happiness. To support this, we take the concept of *Willkür* (choice or elective will) as central and divide duties relating to the agent into two levels: (i) a direct or superior level and (ii) an indirect or subordinate level. The first level concerns maxims that are universalizable in the Kantian sense – that is, maxims that could be willed by all rational beings in any circumstance. From this level derive the perfect and imperfect duties of both right and virtue. At the indirect and subordinate level, there exists a single duty to pursue one's own happiness, whose maxims are not in themselves universalizable. This level is entirely conditioned by the superior level. Therefore, the agent may legitimately pursue their own happiness, provided that doing so does not violate the higher-level duties. In other words, this subordinate duty is only expressed when it does not contradict the direct duties. Finally, this expanded positive model allows for a more substantive distinction between the ideal of happiness (practical good) and immediate pleasure (agreeable).

At this point, a highly relevant issue arises. Once we accept the premise that the CI is the sole principle of practical reason, two possible paths emerge: (i) we may consider that every action is moral (at some level) and therefore there is no space for rationality outside morality (even within actions concerning one's own happiness); or (ii) we may consider that not all actions falling under the CI's domain are moral. If we adopt (i), then we are led to the conclusion that non-moral irrational actions do not exist, since every action is, in some sense, moral. In this case, prudentially irrational actions would still exist, but they would be considered as belonging to the domain of morality rather than to a non-moral domain. On the other hand, if we adopt (ii), then the instrumental principle that guides, for instance, the pursuit of happiness could be seen as a principle grounded in an unconditional principle that nevertheless lies beyond the sphere of morality. In that case, there would be norms of reason that are non-moral (such as, perhaps, the pursuit of one's own happiness).

Nevertheless, I believe that both (i) and (ii) would be highly controversial from a Kantian perspective and would require abandoning key components of his philosophy. Both positions

fail to capture the strict correlation Kant establishes between free action and morality. Kant is particularly concerned with demonstrating that there is a necessary correlation between morally relevant actions and free actions, as we have previously elaborated. It would be far more plausible, from a Kantian standpoint – contrary to what (i) maintains – to recognize the existence of a broad set of non-moral actions, namely, those determined mechanically by nature and which do not contradict the moral law. Choosing whether to buy a house or an apartment is hardly a non-action, but it is also difficult to regard it as a moral action. Moreover, to defend (ii), one would need to argue that every action is, in fact, a free action. This would entail a philosophical structure quite distinct from Kant's, since freedom is understood by him as the *ratio essendi* only of the moral law and, therefore, only of morally relevant actions (KpV AA 05:04). In such a model, *Willkür* would need to possess a spontaneity so extensive that it would surpass even the bounds of morality itself.

5. Concluding remarks

We now enter particularly challenging terrain for the development of a robust theory of irrationality grounded in Kant's philosophy. If we align with Timmerman's position, instrumental irrationality does not exist, and all forms of irrationality must be understood as fundamentally moral. If we adopt Korsgaard's view, we face two problematic alternatives: either every action is regarded as moral – resulting in an inflation of the moral domain – or the essential link between free will and the moral law is undermined. In both scenarios, the synthetic-analytic distinction collapses. The central challenge for those attempting to construct a Kantian account of instrumental irrationality is therefore to explain how instrumental rationality can be possible without reducing it to the moral domain, while simultaneously preserving the intrinsic connection between the moral law and freedom. Put differently, the task is to formulate a theory that satisfies three major conditions: (i) reason exercises a decisive and authoritative role in determining and acting upon instrumental ends; (ii) this form of reasoning is not reducible to morality, and thus does not merely dissolve hypothetical imperatives into components of the categorical imperative; and (iii) this free end-directed agency occurs without presupposing a spontaneity of reason that transcends the boundaries of morality itself. Condition (i) is rejected by Timmerman, whereas conditions (ii) and (iii) remain unfulfilled within Korsgaard's account.

Bibliographic References

HILL, T E. Jr. 2013 Kantian autonomy and contemporary ideas of autonomy. In: SENSEN, O. *Kant on Moral Autonomy*. Cambridge University Press.

KANT, I. 1996. An Answer to the Question: What Is Enlightenment? Trans. James Schmidt. In: SCHMIDT, J. (Ed.). *What is Enlightenment?*: eighteenth-century answers and twentieth-century questions. London: University of California Press.

KANT, I. 2006. *Anthropology from a Pragmatic Point of View*. Trans. Robert B. Louden. Cambridge: Cambridge University Press.

KANT, I. 1987. *Critique of Power of Judgment*. Trans. Werner S. Pluhar. Indianopolis: Hackett Publishing Company.

KANT, I. 1997. *Critique of Practical Reason*. Trans. Mary J. Gregor. Cambridge: Cambridge University Press.

KANT, I. 1998. *Critique of Pure Reason*. Trans. Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.

KANT, I. 2002. *Groundwork for Metaphysics of Morals*. Trans. Allen W. Wood. New York: Yale University Press.

KANT, I. 1900-. *Kants gesammelte Schriften*. Berlin & New York: Walter de Gruyter.

KANT, I. 1991. *The Metaphysics of Morals*. Trans. Mary J. Gregor. New York: Cambridge University Press.

KANT, I. 1996. What does it mean to orient oneself in thinking? Trans. Allen W. Wood. In: WOOD, A. W.; DI GIOVANNI, G. (Orgs.). *Religion and Rational Theology*. Cambridge: Cambridge University Press.

KLEIN, J. T. 2023. Enlightenment as the normative principle of social rationality. *Studia Kantiana*, Curitiba, v. 21, n. 1, p. 99–117. DOI: 10.5380/sk.v21i1.91982. Disponível em: https://revistas.ufpr.br/studiakantiana/article/view/91982. Acesso em: 31 jul. 2025.

KOHL, M. 2018. Kant's Critique of Instrumental Reason. *Pacific Philosophical Quarterly*, v. 99, p. 489–516.

KORSGAARD, C. M. 2008. *The Constitution of Agency*: Essays on Practical Reason and Moral Psychology. Oxford: Oxford University Press.

KORSGAARD, C. M. et al. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

MARTÍNEZ, L. 2023. The Kantian view of dark representations and their function in practical life, according to the anthropological notes of the Critical Period. *Studia Kantiana*, Curitiba, v. 21, n. 1, p. 49–59. DOI: 10.5380/sk.v21i1.91540. Disponível em: https://revistas.ufpr.br/studiakantiana/article/view/91540. Acesso em: 31 jul. 2025.

MERLE, J.-C. 2023. Action irrationality, systemic practical irrationality, and the remedy in Kant. *Studia Kantiana*, Curitiba, v. 21, n. 1, p. 9–18. DOI: 10.5380/sk.v21i1.91472. Disponível em: https://revistas.ufpr.br/studiakantiana/article/view/91472. Acesso em: 31 jul. 2025.

SCHOPENHAUER, A. 1960. *Essay on the Freedom of the Will*. Trans. Konstantin Kolenda. New York: The Liberal Arts Press.

TIMMERMAN, J. 2022. *Kant's Will at Crossroads*: An Essay on the Failings of Practical Rationality. New York: Oxford University Press.

Lies and fake news: a Kantian approach

Mentiras e fake news: uma interpretação kantiana

Karine Cristine de Souza Barboza
Universidade Federal do Paraná (UFPR)/Universität Vechta
karinesouzabarboza@gmail.com

Abstract: This article explores the contributions of Kantian philosophy to contemporary debates on information disorders, with a focus on fake news. It begins with Kant's definition of lying, emphasizing the conditions of declaration, untruthfulness, and intentionality as fundamental criteria for distinguishing misinformation, disinformation, and malinformation. The article argues that, unlike approaches centered on the medium of dissemination, these Kantian criteria offer a more precise delimitation of the concept of fake news. It then develops a critical analysis of Kant's definition of lying within the framework of Doctrine of Right, highlighting the distinction between private law, which protects individual rights, and public law, which safeguards the rights of humanity and social trust—the foundation of the civil pact. The article concludes that the integration of these criteria provides a robust theoretical framework for critically understanding fake news and related information disturbances, contributing to ethical and legal discussions on legitimacy and responsibility in the public sphere.

Keywords: declaration; fake news; information disorders; intentionality; lie; untruthfulness.

Resumo: Este artigo examina as contribuições da filosofia kantiana para o debate atual sobre os distúrbios informacionais, especialmente as fake news. Parte da definição kantiana de mentira, destacando as condições de declaração, inveracidade e intencionalidade como critérios fundamentais para distinguir misinformation, disinformation e malinformation. Defende que, ao contrário das abordagens centradas nos meios de veiculação, esses critérios kantianos oferecem uma delimitação mais precisa do conceito de fake news. Em seguida, desenvolve uma análise crítica da definição kantiana de mentira na Doutrina do Direito, ressaltando a distinção entre direito privado, focado na proteção dos direitos individuais, e direito público, que protege o direito da humanidade e a confiança social, base do pacto civil. Conclui que a articulação desses critérios fornece um referencial teórico sólido para compreender criticamente as fake news e demais distúrbios informacionais, contribuindo para debates éticos e jurídicos sobre legitimidade e responsabilidade na esfera pública.

Palavras-chave: declaração; distúrbios informacionais; fake news; intencionalidade; mentira; inveracidade.

Recebido em 04 de agosto de 2025. Aceito em 29 de setembro de 2025.

1. Introduction

Although Kant does not explicitly offer a specific or extensive discussion of the concept of fake news, the centrality of the theme of lying in his philosophy is undeniable. In the *Groundwork of the Metaphysics of Morals* (GMS), Kant begins with the imperative "thou shalt not lie" from which he explores the scope of the moral duty (GMS, AA, 04: 389 | Kant, 1998a, p. 2-3)[1]. Throughout his writings, the immorality of lying is chiefly examined through the example of false promises. In the *Metaphysics of Morals* (MS), Kant revisits and expands his analysis of lying, characterizing it as "the greatest violation of a human being's duty to himself regarded merely as a moral being" (MS, AA 06: 429 | Kant, 1996, p. 182). There, he distinguishes between external and internal lying, emphasizing the precedence of the former. According to Kant,

> Such insincerity in his declarations [...] still deserves the strongest censure [...] that the ill of untruthfulness spreads into his relations with other human beings as well, once the highest principle of truthfulness has been violated (MS, AA 06: 430-31 | Kant, 1996, p. 183).

As in the *Metaphysics*, Kant also emphasizes the gravity of internal lying in *Religion within the Boundaries of Mere Reason* (RGV). In this context, he addresses it through the lens of untruthfulness in relation to one's own belief in God. Moreover, he defines lying as a form of moral perversity (RGV, AA 06: 38 | Kant, 1998b, p. 60). It is in *Toward Perpetual Peace* (ZeF), however, that we find a discussion of lying in the political context—the very context in which fake news is most commonly debated.

Moreover, despite the editorial and interpretative complexities that characterize the *Lectures on Ethics*, it is possible to identify, in a lecture specifically dedicated to the topic of lying, a passage in which Kant directly addresses the issue of spreading false reports. As the philosopher explains:

> If a man publishes a false report, he thereby does no wrong to anyone in particular, but offends against mankind, for if that were to become general, the human craving for knowledge would be thwarted; apart from speculation, I have only two ways of enlarging my store of information: by experience, and by testimony. But now since I cannot experience everything myself, if the reports of others were to be false tidings, the desire for knowledge could not be satisfied. A *mendacium* is thus a *falsiloquium in praejudicium humanitatis*, even when it is not also in violation of any particular *jus quaesitum* of another. In law a *mendacium* is a *falsiloquium in praejudicium alterius*, and cannot be anything else there, but from the moral viewpoint it is a *falsiloquium in praejudicium humanitatis* (V-Mo/Collins, 27: 447-448 | Kant, 1997, p. 203-204)[2].

The dissemination of false reports, constituting a breach of the fundamental *pactum sociale* described in this passage, finds substantial support in Kant's work, particularly in the *Metaphysics*. This correlation demonstrates that Kant not only acknowledged the far-reaching political and social implications of lying, but also emphasized the importance not merely of truthfulness, but equally of mutual trust among agents as an indispensable condition for both the construction of knowledge and the possibility of social life. Although Kant did not develop a systematic typology of the various forms of falsehood dissemination, this article argues that his concept of lying offers

[1] All references to Kant's works follow the Akademie-Ausgabe (AA) edition (Kant, 1900) accompanied by the corresponding translation.

[2] Although thematically related to this passage from the *Lectures on Ethics*, it is important to note that in the original German text Kant employs the term "*falsche Nachrichten*" (V-Mo/Collins, 27:447–448). In the Cambridge University Press English edition, this expression is rendered as "false report," preserving its broad original sense (Kant, 1997, p. 203). In the Brazilian Portuguese edition published by Unesp (Kant, 2018, p. 300), however, it appears as "notícias falsas" ("fake news"). It should be emphasized that in Kant's context, <*Nachrichten*> is not restricted to the modern journalistic meaning of "news," but rather denotes reports or information in a more general sense. Consequently, translations such as "false reports" or "false information" are more faithful to the original usage, avoiding the semantic narrowing that the contemporary term "fake news" may entail.

significant theoretical resources for the critical analysis and understanding of the contemporary phenomenon of fake news.

Accordingly, the structure of this article has been designed to ensure conceptual clarity and analytical coherence in addressing information disorders through the lens of Kant's practical philosophy. The article begins by proposing a precise delineation of information disorders, specifically disinformation, misinformation, and malinformation. Building on this foundation, it analyzes the threefold distinction found in the specialized literature in light of the conditions established in Kant's definition of a lie. This initial step is justified by the need to avoid terminological ambiguity and to ensure the appropriate applicability of Kantian concepts to the study of information disorders. Subsequently, the article turns to an examination of lying and fake news within the context of Kant's Doctrine of Right, highlighting that such practices constitute significant violations from a juridical perspective, particularly when they undermine public trust and the foundations of the social contract. By adopting this trajectory—from conceptual analysis to normative legal reflection—the study aims to articulate Kantian thought with the contemporary issue of information disorders, providing a robust and critical foundation for the ethical and legal understanding of fake news.

2. Fake news and information disorders

The dissemination and popularization of the term "fake news" gained significant prominence starting in 2016, particularly due to the U.S. presidential elections, when then-candidate Donald Trump accused certain media outlets of spreading false information and being untrustworthy[3]. Previously, the concept of fake news was predominantly associated with two main types of content: "The meaning of the term now ranges from fabricated news circulated via social media to a polemic umbrella term meant to discredit 'legacy' news media" (Quandt *et al.*, 2019, p. 1).

In contemporary discussions, however, the term has become highly ambiguous, due both to the lack of a precise definition and the challenges involved in identifying and characterizing the various modes of dissemination of false information. This concern is echoed by Kant, who holds that lies—particularly false reports—breach a fundamental social pact and engender distrust not only within information channels but also systemic distrust that permeates interpersonal relations, the relationship between citizens and the State, and relations among States, thereby undermining all contracts and, consequently, the very foundation of law.

Regarding the semantic ambiguity of the term fake news, there is a noticeable diversity of approaches and interpretations in the specialized literature. While some authors adopt a broad perspective, others advocate for more restrictive conceptual delimitations, highlighting the complexity and multiplicity of phenomena encompassed by this terminology. Nielsen and Graves (2017, p. 5) report that audiences tend to adopt a broad understanding of the term fake news, encompassing any form of dissemination of false information—including satire, fabricated news, poor journalism, propaganda, and more—regardless of the intent to deceive. By contrast, authors such as Alcott and Gentzkow (2017, p. 213), Lazer *et al.* (2018, p. 1094), Himma-Kadakas (2017), and Gelfert (2018, p. 84) advocate for a more specific use of the term, defining fake news as entirely fabricated content that mimics the form of traditional news, but does not adhere to journalistic processes for verifying the accuracy of information. This group of authors agrees that

[3] For a brief overview of the history of fake news, see Posetti and Matthews (2018).

fake news typifies news presented in the format of traditional media. Gelfert emphasizes that the creation of fake news carries the intention to deceive and is, as the author states, "misleading *by design*" (Gelfert, 2018, p. 84).

Due to the polysemy of the term, the European Commission (Lesher; Pawelec; Desar, 2022, p. 10) advocates abandoning the term fake news, whereas authors such as Lazer *et al*. defend its use "because of its value as a scientific construct, and because its political salience draws attention to an important subject" (Lazer *et al*., 2018, p. 1094).

Despite the divergences surrounding the definition of the term, there is a growing consensus in the specialized literature regarding the importance of distinguishing different forms of false or distorted information, often grouped under the label of "information disorder". The Canadian Centre for Cyber Security (CCCS, 2024) and the European Commission (Lesher; Pawelec; Desar, 2022) all propose a classification into three categories: disinformation, misinformation, and malinformation.

Disinformation refers to false or misleading information that is deliberately created with the intent to manipulate public opinion, generate confusion, obtain political or economic advantages, or cause harm to individuals, institutions, or social groups. The central element of this category is the intention to deceive and produce negative consequences.

Misinformation, in turn, refers to the dissemination of false, inaccurate, or misleading information without deliberate intent to deceive. It often results from errors, misunderstandings, or naïve reliance on unverified sources, and is frequently shared by individuals who believe they are accurately informing others.

Finally, malinformation—or contextual misinformation—involves the use of truthful information that is presented out of context, distorted, or selectively manipulated with the intent to deceive, defame, or cause harm. Typical cases include the disclosure of private messages or genuine data obtained illicitly, as well as the use of accurate content to fuel hate speech or undermine someone's reputation.

In brief, both disinformation and malinformation involve a creator who is aware of the falsity of the information and intends to deceive. These are deliberate acts, unlike misinformation, which results from error or misunderstanding. In the case of misinformation, the spreader is unaware of the falsehood, is not the originator of the content, and lacks any intention to deceive.

In addition to these three distinctions, the European Commission also differentiates between propaganda and satire, thereby broadening the scope of the discussion on forms of misinformation. Regarding the specific placement of fake news within this taxonomy, authors such as Lazer *et al.* (2018) note a significant overlap with the category of disinformation, since fake news is generally fabricated with the deliberate intent to deceive. Distinctly, the European Commission (Lesher; Pawelec; Desar, 2022) following the interpretative framework proposed by Zimdars and McLeod (2020), explicitly classifies fake news as a subtype of disinformation.

Situated within this debate, Kant's definition of lying anticipates and grounds fundamental distinctions essential for understanding the phenomenon of fake news. I will argue below that the concept of fake news aligns closely with Kant's notion of lying, particularly with regard to the conditions of declaration, untruthfulness, and intentionality. These conditions are more pertinent

for delimiting the concept of fake news than the focus on journalistic design, as advocated by authors such as Gelfert (2018, p. 84). The relevant point is that the information is presented within a context of seriousness and veracity, rather than being limited to the format of traditional journalistic media. Although these broader criteria expand the range of cases considered as fake news, they highlight the crucial issue at stake in understanding the manipulation of false information, namely that such information is disseminated within a context that authorizes the receiver's trust both in the veracity of the message and in the sender's commitment to the truthfulness of what is communicated. Therefore, based on a Kantian philosophical framework, both fake news disseminated with deliberate intent to deceive by politicians and state actors, as well as those spread through social media and digital applications, fall within the concept of fake news, not being limited solely to information conveyed through traditional journalistic media formats.

Kant does not restrict false reports to the traditional journalistic format. Rather, he conceives them within a broader spectrum that includes cases such as defamation, false contracts and promises, as well as the employment of lying as a strategy in political practice. Thus, the Kantian definition of lying aligns closely with the broader concept of disinformation, understood as the intentional dissemination of factually false information with the deliberate purpose of deceiving. As will be detailed in the following section, the condition of declaration is fundamental: in order for a lie to occur, the false information must presuppose a declaration of honesty, either explicitly expressed by the communicator or implicitly inferred from the context. This declaration grants the receiver the legitimate right to take the information seriously and to assume that the communicator believes in what is being conveyed. For this reason, based solely on the condition of declaration, the Kantian concept of lying encompasses the tripartite distinction of contemporary information disorders—disinformation, malinformation, and misinformation—as all involve an explicit or implicit commitment by the communicator to the truthfulness of the information or the context in which it is conveyed. However, although the Kantian definition accommodates disinformation (and thus the specificity of fake news) and malinformation, misinformation does not fully fit, as it fails to satisfy the condition of intentionality: in this case, the spreader of misinformation does not have the deliberate intention to deceive but rather shares an error.

The distinction between disinformation and malinformation (or contextual deception) closely parallels the differentiation between lying and paltering. Unlike lying, in paltering the agent is truthful regarding the content of the declaration but deceives by manipulating the context in which the information is conveyed, thereby leading the interlocutor to a mistaken interpretation. This linguistic dynamic has been conceptualized by various authors under different terms: as paltering (Schauer & Zeckhauser, 2009 cited in Mahon, 2015, p. 6); as a deliberate attempt to deceive without resorting to outright falsehood (Saul, 2012); or as a form of linguistic fraud that, while misleading, does not technically constitute lying (Fried, 1978, p. 68). Paltering refers to situations in which the interlocutor is misled through truthful statements combined with the strategic omission of relevant information. It constitutes a form of evasion of sincerity, wherein the speaker, by selecting convenient true propositions, leads the listener to believe they have received a clear and direct answer to the question posed—even though this is not the case. This rhetorical device also frequently appears, for example, in cases of fake news, where real facts are presented out of context, causing the message recipient to infer meanings that were not present in the reported event. This type of malinformation can be illustrated by the case of Menno, presented

in the *Lectures on Ethics*: an example of deliberate ambiguity that simulates a direct answer while avoiding formal lying. As Kant states:

> When Menno, the founder of the sect, was due to be arrested, and escaped by mail coach, the arrestwarrant arrived first, at one of the stages, and the postmaster asked each of the passengers if Menno was on the coach. Instead of lying, that he was not on board, he asked his companion if it was being asked whether Menno was on board; but since the latter did not know Menno, he remained undetected (V-MS/Vigil, AA 27: 702 | Kant, 1997, p. 428).

Kant also points out that, in certain social situations, it is possible to avoid giving a direct answer to a question by means of an unexpected deviation in the conversation, thereby keeping uncertain what the speaker's true or apparent judgment is: "[...] by an unexpected turn of the conversation to divert the others in a direction where it remains doubtful what my true or ostensible judgement will be" (V-MS/Vigil, AA 27: 701 | Kant, 1997, p. 427). This rhetorical device, as described by Kant, constitutes a form of moral ambiguity employed with the intent to deceive. He explains that "*aequivocatio* is permitted, in order to reduce the other to silence and get rid of him, so that he shall no longer try to extract the truth from us, once he sees that we cannot give it to him, and do not wish to tell him a lie" (V-Mo/Collins, AA 27: 449 | Kant, 1997, p. 204–5). As Kant indicates, this kind of subterfuge, known as paltering, does not constitute a lie, as it does not involve the untruthfulness of the information itself: although there is contextual untruthfulness, the information is not factually false.

In this context, Sandel (2012) explores the implications of Kant's definition of lying through his analysis of Bill Clinton's statements during the sex scandal that undermined Clinton's popularity and influenced the impeachment proceedings. After Clinton's romantic involvement with a White House staffer was confirmed, his lawyer argued that the president had misled the public. However, he maintained that Clinton had not technically lied when he claimed not to have had sexual relations with the staffer, since their interaction did not fall under the dictionary definition of "sex". In his provocatively titled essay *Would Kant have defended Bill Clinton*?, Sandel suggests that Kant's conception of lying is significantly narrower than commonly assumed, such that evasive statements—even if technically true, as in the case of paltering—might, under this definition, be considered morally permissible (Sandel, 2012, p. 134).

It is important to note, however, that Kant treats paltering as a strategy employed in situations of social inconvenience, aimed at avoiding embarrassing disclosures. As we will see later in the Doctrine of Right, Kant does not limit legal responsibility to cases of outright lying, but extends it to any form of deception that undermines the freedom of another's choice. In this sense, even within legal contexts, there would be no room for the legitimacy of malinformation, even if it relies on the isolated truthfulness of certain facts, insofar as it distorts the context of their presentation and thereby misleads the recipient.

In what follows, I present the Kantian definition of a lie, elaborating on the conditions of declaration, untruthfulness, and intentionality that constitute it, and examining how these elements relate to the tripartite distinction among types of information disorder.

3. A conceptual framework for fake news grounded in Kant's definition of a lie

Although Kant does not present an exhaustive analysis of the concept of lying, §9 of the Doctrine of Virtue constitutes the primary reference for its definition[4]. Kant defines lying as the "[...] communication of one's thoughts to someone through words that yet (intentionally) contain the contrary of what the speaker thinks" (MS, AA 06: 430 | Kant, 1996, p. 182)[5]. Kant's stylistic choice to place the term "intentionally" <*absichtlich*> in parentheses within the definition of lying, beyond a mere rhetorical device, indicates that intentionality, although not syntactically integrated into the main structure, is nonetheless presupposed as a necessary condition of lying. In this regard, the adverb *absichtlich* performs a dual semantic function: on the one hand, it indicates that lying is a deliberate choice by the agent, rather than a mere mistake; on the other hand, it points to the duplicity of the liar, whose intention to deceive is concealed beneath a pretense of truthfulness. As Kant proceeds, lying is the "[...] renunciation by the speaker of his personality, and such a speaker is a mere deceptive appearance of a human being, not a human being himself". Thus, beyond the intentional expression of the opposite of what one thinks—that is, a conscious propositional untruth—Kant includes in the definition of lying the condition of intentionality, understood as the duplicity of the liar who conceals their true intention.

Kant's definition of lying, grounded in the condition of intent to deceive, encompasses the concepts of disinformation and contextual deception (or malinformation). However, it is important to note that, although the aforementioned tripartite distinction constitutes the prevailing conceptual framework in the study of information manipulation in digital environments, Kant's concept of lying is not limited to the medium through which falsehoods are disseminated. Kant does not confine the concept of lying to oral or written communication, nor does he restrict it to a deceptive interaction between two individuals (one-to-one). Moreover, the act of lying is not limited to speech and can be conveyed through other means such as letters, magazines, television commercials, among others (Mahon, 2015, p. 8).

Consequently, Kant's definition of lying establishes untruthfulness as a fundamental criterion, without necessarily requiring falsehood. As Mahon (2015, p. 5) clarifies, untruthfulness presupposes that the agent utters a declaration he believes to be false, implying the subjective awareness that his proposition does not correspond to reality. In contrast, the condition of falsehood entails that the statement is objectively false. Although contemporary distinctions regarding forms of misinformation predominantly use falsehood as a criterion, it is important to emphasize that for Kant it suffices that the agent intends to deceive through untruthfulness. As will be demonstrated below, when specifically addressing false reports, Kant clearly applies the criterion of falsehood to evaluate the content of the declaration, while employing the criterion of untruthfulness to judge the moral agent's intention.

Related to this distinction between untruthfulness and falsehood, the criterion of intentionality in the definition of lying serves as a basis for differentiating between lying and error, much like the distinction between disinformation and misinformation. According to the philosopher,

[4] For a detailed study of the conditions of the concept of lying in Kant, see Barboza and Klein (2024), upon which this section draws.

[5] In the original text: "[...] die doch das Gegenteil von dem (absichtlich) enthalten, was der Sprechende dabei denkt [...]" (MS, AA 06: 429).

> One cannot always stand by the truth of what one says to oneself or to another (for one can be mistaken); however, one can and must stand by the truthfulness of one's declaration or confession, because one has immediate consciousness of this. For in the first instance we compare what we say with the object in a logical judgment (through the understanding), whereas in the second instance, where we declare what we hold as true, we compare what we say with the subject (before conscience). Were we to make our declaration with respect to the former without being conscious of the latter, then we lie, since we pretend something else than what we are conscious of (MpVT, AA 08: 267-268 | Kant, 1998b, p. 27).

In this passage, Kant emphasizes that the decisive moral standard for evaluating declarations is not their conformity with the facts, but rather their truthfulness—understood as the correspondence between what is declared and what is believed. This stems from the fact that, while error concerning objects external to consciousness is always possible, the same does not apply to internal content—that is, to what we are subjectively aware of. Conceptually, therefore, truthfulness and untruthfulness refer to the relationship between a declaration and the agent's belief, whereas truth and falsehood pertain to the correspondence between the declaration and cognitive content.

Another important point, clarified by Wood (Wood, 2008, p. 240), is that lying must be understood as a declaration (*Aussage, Deklaration, declaratio*), that is, a form of expression that grants the listener the right to presume that what is said corresponds to what is believed. A declaration, therefore, entails a commitment to truthfulness, as it occurs in contexts in which the interlocutor is entitled to rely on the conformity between speech and the speaker's internal conviction—which makes the speaker morally responsible if they declare something they know to be false (Wood, 2008, p. 241).

The commitment to truthfulness may be explicitly stated by the speaker or, as is often the case, implied by the context itself. In normative settings, such as ethical or legal domains, the requirement of honesty is presumed, since the context authorizes interlocutors to expect that what is said corresponds to what is believed. However, there are discursive situations in which this requirement is suspended, either explicitly or contextually—as in the case of satire or fictional works. In such cases, the utterance is not subject to the implication of truthfulness, and the listener is not expected to take what is said at face value. As Kant observes, when a speaker signals that their words are not to be taken seriously—whether by stating so directly or through the circumstances in which the statement is made—one cannot say that a lie has been told, even if someone is misled as a result. In such a case, the deception does not stem from a deliberate intention to deceive (MS, AA 06: 238 | Kant, 1996, p. 30). Although such expressions may, in fact, lead to misunderstanding, the absence of an intention to deceive, combined with the suspension of declarative commitment, exempts the speaker from moral responsibility for lying—even if they may, in certain cases, still be held accountable for the harm caused by the misunderstanding, regardless of whether it was grounded in a lie.

The European Commission (Lesher; Pawelec; Desar, 2022, p. 9) defines satire as a form of social and political critique that uses humor and exaggeration to address relevant societal issues. However, a disinformation problem arises when the original satirical context is lost, often due to mass sharing. This loss of context may lead the reader to interpret the content as a literal assertion. Such a situation is particularly problematic for the declarative condition in the Kantian definition of lying, as the recipient is left uncertain about how to interpret the information. However, even when the liar takes advantage of such ambiguity to deceive, the Kantian definition of lying does not exempt them from the harm resulting from their action. The Kantian definition of lying

includes the condition of untruthfulness, which is grounded in the agent's intention rather than the falsehood of the information. Kant states that the liar "makes himself contemptible in his own eyes and violates the dignity of humanity in his own person" (MS, AA 06: 429 | Kant, 1996, p. 182). In the legal sphere, although the ambiguity of the context may hinder the proof of intent to deceive, the agent can still be held liable for the damages caused by their deception. Thus, according to the tripartite distinction of information disorders, the loss of the satirical context and contextual ambiguity can be understood as a use of contextual deception or malinformation by the spreader, which may generate misinformation in the receiver—these points will be further developed in the following section, dedicated to lying in the Doctrine of Right.

As initially argued, Kant's concept of lying provides central conceptual elements for understanding the contemporary phenomenon of fake news. Thus, even the emphasis on the condition of declaration—fundamental in Kant's definition—indicates that the study of misinformation should not be confined to the traditional journalistic context. Restricting the analysis to this specific domain risks overlooking other equally relevant communicative contexts. The criterion of declaration, as employed by Kant, is more appropriately used to delineate cases of information disorders: what matters is not the medium used, but whether the interlocutor is entitled to take the received information seriously, presuming that the communicator is committed to the truthfulness of what is conveyed. As highlighted by various contemporary studies, this type of communicative commitment manifests across multiple mediation forms that go beyond magazines and newspapers, including social networks, messaging applications, and other digital formats of information circulation.

4. Fake news and lying in the Doctrine of Right

In the passage where Kant asserts that "if a man publishes a false report, he thereby does no wrong to anyone in particular, but offends against mankind", the philosopher revisits the distinction between the public and private spheres of law in the Doctrine of Right (V-Mo/Collins, AA 27: 448 | Kant, 1997, p. 203-204). In private law, lying is narrowly defined as a declaration contrary to what one thinks, uttered with the intention to deceive and resulting in harm to another's free will. In this sense, Kant restricts the legal use of the term "lie" to falsifications that directly infringe upon the rights of another person, as exemplified by the simulation of a contract (MS, AA 06: 238 | Kant, 1996, p. 30). He clarifies that the jurist recognizes and applies this definition only when there is a violation of legal duties toward others (*officii juridicorum*), understanding lying in this case as a *falsiloquium dolosum in praejudicium alterius*—that is, a deceitful falsehood to the detriment of another (V-MS/Vigil, AA 27: 604–605 | Kant, 1997, p. 350–351).

Consequently, public law understands *mendacium* more broadly, recognizing it as a violation not only of legal duties between individuals but also as an infringement of the right of humanity. Kant observes that even when a lie does not cause direct harm to anyone, it still constitutes a serious breach of duty in general, as it undermines the credibility of declarations. As he states,

> I do wrong to duty in general in a most essential point. That is, as far as in me lies I bring it about that statements (declarations) in general find no credence, and hence also that all rights based on contracts become void and lose their force, and this is a wrong done to mankind in general (VRML, AA 08: 426 | Kant, 1993, p. 64).

Thus, Kant concludes that lying inevitably causes harm: if not to an individual in particular, then at least to humanity as a whole, since it corrupts the very source of law. In this context, trust in the truthfulness of declarations is a structural and indispensable element for legal stability

and legitimacy. Therefore, whether lying in court, in contracts, or disseminating fake news, the offender does not only harm the directly affected party but also weakens the entire legal system based on the credibility of declarations (Wood, 2008, p. 243). For this reason, the conditions of declaration, untruthfulness, and intentionality are crucial for the configuration of criminal or civil infractions, as well as for the proper delimitation of corresponding legal responsibility.

Based on Green's analysis of perjury, for example, it is possible to highlight the importance of distinctions grounded in the criteria present in Kant's conception of lying. As Green emphasizes, the legal treatment of deception varies depending on the social role of the agent—whether an ordinary citizen or a public authority—and on the specific context in which the deception occurs, such as judicial hearings, commercial transactions, police stations, or situations of sexual intimacy (Green, 2019, p. 483). The social role of the individual who deceives, as well as the context in which the deception occurs, reinforce the significance of the declaration condition. For example, a public authority, in the exercise of their duties, is committed to adhering to the procedures established by the norms regulating their civil competence. Moreover, the context in which the declaration is made supports or delineates its declarative character, determining the specific requirements applicable to each situation.

The legal relevance of the untruthfulness condition in lying is particularly evident in cases of perjury. According to Green, statutes concerning perjury have traditionally been interpreted as requiring that the declaration be objectively verifiable with respect to its untruthfulness or falsehood. Declarations based on beliefs or opinions do not constitute perjury, except in exceptional cases where the witness claims to hold a belief or opinion that they, in fact, do not possess—this existence or nonexistence of the belief being a fact susceptible to verification.

The author further highlights the complexity of the evidentiary requirement regarding whether it is necessary to demonstrate that the statement was literally false, or if a "merely misleading" declaration, even if literally true, would suffice to support a conviction. As Green explains, the rule currently followed in most common law jurisdictions requires that, for a witness to be convicted of perjury, they must actually lie under oath, and not merely mislead (Green, 2019, p. 485). In other words, just as in Kant's definition of lying, the distinction between the untruthfulness of the declaration and factual falsity, along with the intention to deceive, are fundamental elements for delimiting the legal offense. Thus, it becomes essential to determine whether there was a declaration contrary to what the speaker believes, characterizing a lie, or whether the deception occurred through other modalities, such as in the case of paltering.

Green illustrates this distinction by revisiting the case of Bill Clinton, cited in Sandel's critique of Kant's definition of lying (2012, p. 134). Despite being charged with perjury by the Independent Counsel, Green contends that Clinton's response about being alone in the Oval Office with the White House intern does not amount to perjury. Green would likely agree with Sandel on this point, stating that "Clinton offered an evasive, non-responsive, and factually true reply to the question posed; but he did not actually lie" (Green, 2019, p. 486). Although Clinton intended to deceive, Green argues that his conviction could not be grounded in the legal definition of perjury: the legal context of Clinton's testimony certainly imposed the declarative condition, and there was intent to deceive; however, by providing an evasive yet factually true response, no falsification of the declaration occurred. It should be noted, however, that neither Green nor Sandel deny Clinton's intent to mislead the Court.

The intentionality condition inherent in Kant's concept of lying is fundamental for analyzing the responsibility for untruthfulness. Kant reaffirms the formal injustice of lying in his treatment of false reports, stating that "every lie is objectionable and deserving of contempt, for once we declare that we are telling the other our thoughts, and fail to do it, we have broken the pactum, and acted contrary to the right of humanity" (V-Mo/Collins, 27: 448 | Kant, 1997, p. 203–204). The broken pact is described by Kant as the second condition of sociability: "but the liar destroys this fellowship, and hence we despise a liar, since the lie makes it impossible for people to derive any benefit from what he has to say" (V-Mo/Collins, 27: 444 | Kant, 1997, p. 201). Truthfulness, as a condition of sociability, requires correspondence between the intention declared in discourse and the agent's actual intention. Hence, there is an intrinsic tension between lying and truthfulness: the former presupposes the intention to deceive, whereas the latter explicitly rejects it. Although Kant repeatedly emphasizes the immorality of lying in various texts, the Doctrine of Right also provides sanctions for other forms of deception, since any falsification of the declared intention violates the legal duty of truthfulness. What makes lying—and, in legal terms, perjury—particularly serious is the deliberate declaration of the opposite of what one believes, directly concealing one's true intention in contexts where there is an explicit duty of truthfulness.

Kant characterizes lying as a formal injustice against the right of humanity also in the *Metaphysics*, denying that false reports constitute a right even in the case of war between states, since such behavior "would make its subjects unfit to be citizens [...] in a word, using such underhanded means as would destroy the trust requisite to establishing a lasting peace in the future" (MS, AA 06: 348 | Kant, 1996, p. 117).

The first relevant point regarding this passage concerns the dissemination of false reports as a harm to the free will of others, such that the falsity of these declarations results in damage both epistemologically and politically. Kant emphasizes the epistemological aspect in the *Lectures on Ethics* when he states that "[...] since I cannot experience everything myself, if the reports of others were to be false tidings, the desire for knowledge could not be satisfied" (V-Mo/Collins, 27: 448 | Kant, 1997, p. 203–204). For Kant, although it is not essential that the ideas expressed be true (since error is possible), the lack of truthfulness would undermine the credibility of discourse, reducing it to a mere collection of assertions without commitment on the part of their authors. This situation would open the door to the unrestricted propagation of unfounded information, thereby weakening standards of responsibility, including within the legal sphere. Therefore, Kant maintains that assuming responsibility for one's own declarations is fundamental for the legitimate exercise of public reason.

Kant further develops the political harms of false reports in *Toward Perpetual Peace*. In addition to establishing falsehood as a moral limit to political prudence (ZeF, AA 08: 370 | Kant, 1917, p. 162), Kant rejects all peace treaties concluded with *reservatio mentalis*, that is, with mental reservations (ZeF, AA 08: 344 | Kant, 2006, p. 68). This prohibition is grounded in the moral imperative to avoid any form of deception in declarations of intent, especially those that concern the end of war. In the philosophical tradition, *reservatio mentalis* denotes a form of dissimulation in which a speaker limits the meaning of their words internally while outwardly expressing them without any restriction. It is a deceptive ambiguity: the speaker utters a declaration that appears complete and truthful, yet secretly adds an internal condition that alters its actual meaning.

For Kant, this kind of dissimulation is morally impermissible in juridical contexts such as treaties because it violates the condition of public truthfulness. Even if it does not involve an explicit lie—since the speaker does not affirm something they believe to be false—it still aims to mislead the other party. As such, *reservatio mentalis* constitutes a species of falsehood that undermines the trust necessary for rightful relations between states. In the context of peace treaties, it conceals the true intention not to maintain peace, thereby nullifying the validity of the agreement and transforming it into a covert instrument of war. Thus, Kant considers treaties made with mental reservation not only illegitimate but also morally equivalent to a falsification, because the communicative act pretends to bind the parties while inwardly intending the opposite.

As Wood argues, Kant was attentive to the political implications of permitting lying: "the issue that appears to have really concerned both Kant and Constant is the duty of politicians and statesmen to be truthful in their official declarations" (Wood, 2008, p. 249)[6].

Kant synthesizes in his response to Constant the arguments against political legitimacy for lying that he had previously developed in *Toward Perpetual Peace.* In this work, Kant presents the principles "Act and justify yourself" (*Fac et excusa*), "If you did it, deny it" (*Si fecisti nega*), and finally "Divide and rule" (*Divide et impera*) (ZeF, AA 08: 374–375 | Kant, 2006, p. 98). Kant describes that under the principle "Act and justify yourself", committed violence is falsely justified after the fact, generating lies regarding the legitimacy of the state's possession. Consequently, Kant describes that under the principle "If you did it, deny it", the politician lies about the mistakes committed, fabricating falsehoods to attribute the errors and illegalities to his subjects. Lastly, Kant revisits the famous principle "Divide and rule", portraying it as a political simulation aimed at fomenting discord among the people in order to deceive them with promises of greater freedom. Kant concludes this passage by stating that "it is true that nowadays nobody is taken in by these political maxims, for they are all familiar to everyone" (ZeF, AA 08: 376 | Kant, 2006, p. 172). In commenting on this passage, Kant criticizes the normalization of political cynicism among great powers, emphasizing that these unjust maxims—such as lying and manipulation—are no longer a source of scandal, since their public acknowledgment causes no shame. What discredits rulers is not the immorality of these principles, but their failure in practice, given that political prestige is measured by the expansion of power rather than by moral legitimacy. This position directly contrasts with Kant's own view that politics must remain subordinate to morality. As he argues in Toward Perpetual Peace, the success of political action can never justify the adoption of immoral maxims, for political conduct must be guided by universal principles of justice and respect for human dignity.

5. Conclusion

This article set out to investigate the complex conceptual issue of fake news through Kant's philosophical framework, aiming to overcome the ambiguities and disputes that permeate contemporary debates on information disorders. Reaffirming the central thesis, the criteria for defining lying in Kant—namely, declaration, untruthfulness, and intentionality—prove essential to deepen the contemporary discussion on information disorders.

[6] For a reconstruction of the dispute between Kant and Constant, as well as a proposed resolution of the moral casuistry, see Klein (2018).

The analytical function of the declaration condition is to define what is formally subject to verification of truth or veracity. That is, it delimits cases in which there is a clear and morally relevant propositional content. The condition of untruthfulness, in turn, allows distinguishing between factual falsity and the untruthfulness of the declaration—that is, between objectively false information and the absence of commitment to the truth. The condition of intentionality, like untruthfulness, enables the differentiation of lying proper from other types of false declarations that do not involve intent, such as errors or unintentional misinformation. This condition is fundamental, for example, in distinguishing disinformation (false and intentional) from misinformation (false, but not necessarily intentional).

In their articulation, these criteria provide a robust and precise conceptual framework that enables a rigorous analysis of the multiple facets of information disorders, distinguishing them clearly and offering solid grounds for ethical and legal analyses. Kant's work presents crucial criteria not only for the definition of information disorders but also addresses these issues within the theory of law, the legal system, and political contexts. Thus, Kant's definition of lying decisively contributes to a critical understanding of fake news, broadening and deepening reflections on the moral and social challenges inherent in information disorders.

The articulation between the criteria for defining fake news and the specificity of information disorders for its analysis within the Doctrine of Right is crucial. This demonstrates that lying, as a violation of the right of humanity and the social pact, goes beyond mere isolated falsehoods or traditional media formats. It constitutes a transgression of the epistemological and political foundations of social coexistence and the legal order. Kant distinguishes between lying that violates individual rights—typical of private law—and lying that affects the right of humanity, undermining social trust and the foundation of the civil pact in public law. Thus, Kantian conditions for the definition of lying allow for capturing decisive elements in the contemporary dynamics of fake news: the relevance of context and the imposed demands, delineated by the condition of declaration; as well as the untruthfulness and intentionality of the liar in declaring the opposite of what they intend. This approach broadens the analytical scope to encompass the multiple channels through which fake news spreads—from social media and digital applications to traditional media—and incorporates the diversity of political and interpersonal environments in which it manifests. In this regard, Kantian reflection not only offers conceptual clarity but also highlights the profoundly harmful nature of fake news. Consequently, Kantian thought proves to be an indispensable epistemological and normative tool for the critical understanding and confrontation of information disorders.

Bibliographic References

ALLCOTT, H.; GENTZKOW, M. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, v. 31, n. 2, p. 211-236. DOI: 10.1257/jep.31.2.211

BARBOZA, K. C. S; KLEIN, J. T. 2024. O conceito de mentira segundo Kant. *Revista Instante*, v. 6, n. 3, p. 42–174. DOI: 10.29327/2194248.6.3-7

CANADIAN CENTRE FOR CYBER SECURITY. 2024. *How to identify misinformation, disinformation, and malinformation*. Awareness series. Available at: https://www.cyber.gc.ca/en/guidance/how-identify-misinformation-disinformation-and-malinformation-itsap00300#wb-tphp. Accessed in July 2025.

FRIED, C. 1978. *Right and wrong*. Oxford: Harvard University Press.

GELFERT, A. 2018. Fake news: a definition. *Informal Logic*, v. 38, n. 1, p. 84-117. DOI: 10.22329/il.v38i1.5068

GREEN, S. P. 2019. Lying and the law. In: MEIBAUER, J. (Ed.). *The Oxford handbook of lying*. Oxford: Oxford University Press, p. 483-493.

HIMMA-KADAKAS, M. 2017. Alternative facts and fake news entering journalistic content production cycle. *Cosmopolitan Civil Societies*, p. 25-40. DOI: 10.5130/ccs.v9i2.5469

KANT, I. 1900. *Gesammelte Schriften v. 1-24*. Berlin: Preussische Akademie der Wissenschaften; Deutsche Akademie der Wissenschaften zu Berlin (v. 23); Akademie der Wissenschaften zu Göttingen (v. 24).

KANT, I. 1993. *Grounding for the metaphysics of morals with on a supposed right to lie because of philanthropic concern*. Translated by James Ellington. Indianapolis: Hackett Publishing Company.

KANT, I. 1998a. *Groundwork of the metaphysics of morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press.

KANT, I. 1997. *Lectures on ethics*. Translated by Peter Heath. Cambridge: Cambridge University Press.

KANT, I. 2018. *Lições de ética*. Traslated by Bruno Leonardo Cunha and Charles Feldhaus. São Paulo: Editora Unesp Digital.

KANT, I. 1917. *Perpetual peace: a philosophical essay*. Translated by M. Campbell Smith. London: George Allen & Unwin Ltd.

KANT, I. 1998b. *Religion within the boundaries of mere reason and other writings*. Translated by Allen Wood and George Di Giovanni. Cambridge: Cambridge University Press.

KANT, I. 1996. *The metaphysics of morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press.

KANT, I. 2006. *Toward perpetual peace: a philosophical sketch*. Translated by David Colclasure. New Haven: Yale University Press.

KLEIN, J. T. 2018. Kant versus Constant: sobre um suposto direito de mentir. *Studia Kantiana*, v. 16, p. 95-126. DOI: 10.5380/sk.v16i3.89814

LAZER, D. M. J. *et al*. 2018. The science of fake news: addressing fake news requires a multidisciplinary effort. *Science*, v. 359, n. 6380, p. 1094-1096. DOI: 10.1126/science.aao2998

LESHER, M.; PAWELEC, H.; DESAR, A. 2022. Disentangling untruths online: creators, spreaders and how to stop them. *OECD Going Digital Toolkit Notes*, n. 23, p. 1-37.

MAHON, J. 2015. *The definition of lying and deception*. Stanford Encyclopedia of Philosophy. Available at: https://plato.stanford.edu/archives/fall2015/entries/lying-definition/. Accessed in July 2025.

NIELSEN, R. K.; GRAVES, L. 2017. *"News you don't believe"*: audience perspectives on fake news. Oxford: Reuters Institute for the Study of Journalism. Available at: https://reutersinstitute.politics.ox.ac.uk/our-research/news-you-dont-believe-audience-perspectives-fake-news. Accessed in July 2025.

POSETTI, J.; MATTHEWS, A. 2018. *A short guide to the history of 'fake news' and disinformation*: a learning module for journalists and journalism educators. ICFJ. Available at: https://www.icfj.org/news/short-guide-history-fake-news-and-disinformation-new-icfj-learning-module. Accessed in July 2025.

QUANDT, T. *et al*. 2019. Fake news. In: VOS, T. P. *et al*. (ed.). *The international encyclopedia of journalism studies*. Hoboken: John Wiley & Sons.

SANDEL, M. 2012. *Justice*: what's the right thing to do? New York: Farrar, Straus and Giroux.

SAUL, J. M. 2000. Did Clinton say something false? *Analysis*, v. 60, n. 3, p. 255-257. DOI: 10.1111/1467-8284.00235

SCHAUER, F.; ZECKHAUSER, R. 2009. Paltering. In: HARRINGTON, B. (Ed.). *Deception*: from ancient empires to internet dating. Stanford: Stanford University Press, p. 38-54.

WOOD, A. 2008. *Kantian ethics*. Cambridge: Cambridge University Press.

ZIMDARS, M.; McLEOD, K. 2020. *Fake news*: understanding media and misinformation in the digital age. Cambridge: MIT Press.

Habermas on social irrationality

Habermas e a irracionalidade social

Cristina Foroni Consani[1]
Universidade Federal do Paraná (UFPR)
cristina.foroni@ufpr.br

Abstract: This paper seeks to identify what constitutes social irrationality within the Habermasian framework. The discussion is structured in two parts: first, I delineate the definitions of rationality and irrationality in Habermas's work. I argue that while social rationality and irrationality are most prominent within the intersubjective world and the sphere of communicative rationality, they are also identifiable within the domains of epistemic and teleological rationality. Second, I examine the extent to which Habermas's concept of rationality offers answers to practical issues arising from social irrationality.

Keywords: Habermas; communicativ rationality; epistemic rationality; rationality; social irrationality; teleological rationality.

Resumo: Este artigo tem como objetivo identificar o que pode ser considerado como irracionalidade social na obra de Habermas. Este tema será abordado em dois momentos: primeiro, apresentarei as definições de racionalidade e irracionalidade na obra de Habermas. Argumentarei que, embora a racionalidade e a irracionalidade sejam mais evidentes no mundo intersubjetivo e no domínio da racionalidade comunicativa, elas também podem ser identificada no domínio da racionalidade epistêmica e teleológica. Em segundo lugar, analisarei até que ponto o conceito de racionalidade na obra de Habermas pode fornecer respostas para questões práticas relativas à irracionalidade social.

Palavras-chave: Habermas; racionalidade comunicativa; racionalidade epistêmica; racionalidade; irracionalidade social; racionalidade teleológica.

[1] This study was partly financed by National Council for Scientific and Technological Development – CNPq – and partly by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Recebido em 30 de agosto de 2025. Aceito em 19 de janeiro de 2026.

In recent years, several global phenomena have reignited the debate surrounding social irrationality. The COVID-19 pandemic, for instance, offered numerous examples, from the rejection of scientific data—evidenced by anti-vaccine movements and the promotion of unproven medical treatments—to protests that escalated into the vandalism of public infrastructure as a means of opposing social distancing measures. This broader trend encompasses climate change denialism and political movements that undermine the foundations of constitutional democracies through misinformation and unjustified resentment, as exemplified by the 2021 January 6 U.S. Capitol attack and the 2023 January 8 Brasília attacks.

These events represent social phenomena that can be scrutinized through the lens of social irrationality—a concept whose interpretation shifts significantly depending on the specific theory of rationality employed. This paper seeks to identify what constitutes social irrationality within the Habermasian framework. The discussion is structured in two parts: first, I delineate the definitions of rationality and irrationality in Habermas's work. I argue that while social rationality and irrationality are most prominent within the intersubjective world and the sphere of communicative rationality, they are also identifiable within the domains of epistemic and teleological rationality. Second, I examine the extent to which Habermas's concept of rationality offers answers to practical issues arising from social irrationality.

I – Rationality and Social Irrationality

Through the method of rational reconstruction , Habermas aims to uncover elements of an inherent, yet under-explored, rationality within the reproduction of society as a whole (cf. HABERMAS, 2008, p. 24). Given that former religious and metaphysical worldviews have been superseded by modern science and pluralism—conditions that permit only a formal rather than substantive concept of rationality—his theory offers a procedural framework. As Habermas notes: "[t]here is no pure reason that might don linguistic clothing only in the second place. Reason is by its very nature incarnated in contexts of communicative action and in structures of the lifeworld." (cf. HABERMAS, 1990, p. 322) Thus, his analysis centers on the structures of justification[2].

Since the mid-1970s, Habermas has distinguished between two primary forms of rationality: *teleological* (instrumental and strategic) and *communicative*. The former refers to the use of resources, efficiency, and the relationship between means and ends, while the latter concerns the possibility of "identify[ing] and reconstruct[ing] universal conditions of possible understanding" (HABERMAS, 1976, p. 1; 1984, p. 8).

More recently, in *On the Pragmatics of Communication* (1998) and *Truth and Justification* (1999), the philosopher has further specified the roles, scopes, and interrelationships of these types of rationality. To avoid a purely subject-centered approach, Habermas notes in *On the Pragmatics of Communication* that he does not find "the proposal to reduce rationality to a disposition of rational persons promising." (HABERMAS, 1998, p. 308). Instead, he maintains that the predicate 'rational' applies "to refer to *beliefs*, *actions* and *linguistic utterances* because, in the propositional structure of knowledge, in the teleological structure of actions, and in the communicative structure of speech, we come upon *varios roots of rationality*." (HABERMAS, 1998, p.

[2] See also Dutra, 2005, p. 42.

308-309). In this sense, rationality has three distinct roots (*Wurzeln*)[3], namely, epistemic, teleological, and communicative. (cf. HABERMAS, 1998, p. 309; HABERMAS, 2004, p. 105).

Epistemic rationality is defined by the propositional structure of knowledge, consisting of judgments that are subject to being true or false—for instance, the contemporary claim that the Earth is flat is demonstrably false. Within this framework, "knowledge is intrinsically of a linguistic nature" and therefore lies within the scope of justification and criticism. Knowledge recognized as true should not, however, be considered an unconditional truth, as epistemic rationality only recognizes truth in the sense of its justified acceptability in a given context. In this regard, it is emphasized that "the explicit 'knowing what' is bound up implicitly with a 'knowing why' and insofar points toward potential justifications" (HABERMAS, 1998, p. 312).

Irrationality, accordingly, is not synonymous with falsehood. As the philosopher notes, "[w]hoever shares views that turn out to be untrue is not *eo ipso* irrational. Someone is irrational if she puts forward her beliefs dogmatically, clinging to them although she sees that she cannot justify them" (HABERMAS, 1998, p. 312). Epistemic irrationality is thus identified with a failure of justification rather than an error in reasoning. In this sense, the ancient belief in a flat Earth was an understandable error—not an act of irrationality—given the limited information available at the time. Today, however, maintaining that the Earth is flat is irrational because the claim can no longer be justified. This correlation between irrationality and the absence of justification extends to the other roots of rationality as well.

Teleological rationality is associated with intentional action. The rationality of an act is measured by whether the agent achieves a desired result through deliberately chosen and employed means. Within this framework, a successful actor is considered to have acted rationally only if he: "(i) knows why he was successful (...) and if (ii) this knowledge motivates the actor (at least in part) in such a way that he carries out his action for reasons that can at the same time explain its possible success" (HABERMAS, 1998, p. 313-314). Irrationality in this context consists of selecting inappropriate means to achieve desired ends. To illustrate, a student might establish the goal of passing an exam but, instead of studying, relies solely on luck. In doing so, she acts irrationally because such means lack a grounded relationship to the intended outcome. The irrationality lies not in the goal itself, but in the agent's failure to orient her action toward means that can objectively explain or justify the potential for success.

Purposive-rational action also requires reflexivity and adaptation to possible justifications. According to Habermas, "there is a relationship of mutual reference between the rationality of the action and the forum of a discourse which an actor's decisive reasons for making his decision—determined *ex ante*—could be tested" (HABERMAS, 1998, p. 314). In this vein, teleological rationality is argued to be intertwined "with the two other core structures of knowledge and speech" given that "the practical considerations by means of which a rational plan of action is carried out are dependent on the input of reliable information (about expected events in the world, or about the behavior and the intentions of other actors)" (HABERMAS, 1998, p. 314).

Conversely, it is important to clarify that such information is accessible strictly through linguistic representation, focusing exclusively on the goals chosen by the agent based on her per-

[3] Habermas employs the term "roots" (*Wurzeln*) of rationality to describe these distinct origins; however, in some passages he alternately uses the terms "Structure" (*Struktur*) or "core structures" (*Kernstrukturen*) of rationality. See HABERMAS, 2004, pp. 104, 105, 110.

sonal interests—independent of others' concerns or any form of external debate. In this regard, "elementary action-intentions and simple practical inferences, too, are linguistically structured. Just as propositional knowledge is dependent on the use of propositional sentences, so too is intentional action essentially dependent on the use of intentional sentences" (HABERMAS, 1998, p. 314-315).

Communicative rationality, in turn, manifests as the "unifying force of speech oriented toward reaching understanding" (HABERMAS, 1998, p. 315). On this basis, it is argued that:

> We do not call only valid speech acts rational but rather all comprehensible speech acts for which the speaker can take on a *credible* warranty in the given circumstances to the effect that the validity claims raised could, if necessary, be vindicated discursively (HABERMAS, 1998, p. 315-316).

In this context, the rationality of a speech act remains intrinsically linked to its potential justification, as it is through argumentation that implicit validity claims are thematized and examined based on reasons. This inherent rationality is anchored in the internal connection between:

> (a) the conditions that make a speech act valid, (b) the claim raised by the speaker that these conditions are satisfied, and (c) the credibility of the warranty issued by the speaker to the effect that he could, if necessary, discursively vindicate the validity claim (HABERMAS, 1998, p. 316-317).

These validity claims are specified as follows*: truth claims*, which refer to facts within the objective world; *claims to truthfulness*, involving statements that reveal subjective experiences to which the speaker has privileged access; and *claims to the rightness* of norms and commands, which pertain to the search for recognition within an intersubjectively shared social world. Accordingly, within the framework of communicative rationality, irrationality occurs when the requisite justifications for these validity claims are not provided. Based on these definitions, irrationality is fundamentally characterized by the absence of the justification necessary to sustain each respective type of rationality.

In presenting these three roots of rationality, Habermas offers two essential caveats. First, these structures exist on the same level; that is, communicative rationality does not function as an overarching framework, but is rather one of three central structures interwoven by discursive rationality. Second, these structures cannot be conceived in a mentalist fashion, as "epistemic and teleological rationality are not of a prelinguistic nature". (HABERMAS, 1998, p. 309). While the first warning highlights the social character of rationality—insofar as it pertains to intersubjective interactions[4], the second emphasizes that the use of language can be both communicative and non-communicative.

Regarding the first caveat, discursive rationality is defined as a procedural rationality inherent in the practices of justification. It is associated with the capacity to engage in argumentative practices—specifically, the practice of criticizing and justifying problematic claims by providing reasons and arguments. This type of activity is what Habermas calls "discourse"; he employs the term "discursive rationality" to refer to the comprehensive set of competencies a speaker must acquire to participate in argumentation as a reflexive form of communication. This is further clarified in the following passage:

> The rationality of a person is proportionate to his expressing himself rationally and to his ability to give account for his expressions in a reflexive stance. A person expresses himself rationally insofar as he is oriented performatively toward validity claims: we say that he not only behaves rationally but is himself rational if he can give account for his orientation toward validity claims. We also call this kind of rationality *accountability*

[4] See REED; MOORE, 2019, p. 379.

> *(Zurechnungsfähigkeit)*.
> Accountability presupposes a reflected self-relation on the part of the person to what she believes, says, and does; this capacity is entwined with the rational core structures of knowledge, purposive activity, and communication by way of the corresponding self-relations (HABERMAS, 1998, p. 310).

The alignment of rationality with accountability serves as a cornerstone for Habermas's delineation of the scope of his theory. As explained in *Truth and Justification*, rationality is a supposition within contexts of action oriented toward reaching understanding—a supposition "that anyone engaged in communicative action must assume." Under this assumption, "a subject who is acting intentionally is capable, in the right circumstances, of providing a more or less plausible reason for why she did or did not behave or express herself one way rather than another". Consequently "[s]omeone who cannot account for her actions and utterances to others becomes suspect of not having acted reasonably or "accountably" [*zurechnungsfähig*] (HABERMAS, 2003, p. 94).

Habermas draws a parallel between his definition of accountability and Kant's conception of freedom (HABERMAS, 1998, p. 310; 2003, p. 94). This comparison highlights the fundamental differences between communicative rationality and Kantian practical reason. While Kantian theory links freedom to moral standards and purposive rationality, Habermas maintains that accountability is evaluated through the validity claims raised according to the core structures of knowledge, action, and speech. In this sense, accountability "involves more than just practical reason. Accountability consists, rather, in an agent's general ability to orient her action by validity claims". (HABERMAS, 2003, p. 95).

In Kant's philosophy, practical reason establishes universal laws and possesses both a categorical sense of obligation and a transcendental sense of certainty. It also suggests that autonomous action is a possibility, rather than being merely counterfactual. Conversely, rationality within communicative action does not constitute an obligation, even regarding moral or legal conduct. Instead, it delineates what it means to act autonomously. It presupposes that all participants are responsible agents who position themselves based on validity claims. (cf. HABERMAS, 2003, p. 96). In essence, communicative rationality differs from Kantian practical reason by encompassing a broader array of validity claims—incorporating truth and truthfulness alongside rightness. For Habermas, its role is not to dictate norms of conduct directly, but rather to provide the framework for an orientation toward validity claims within processes of argumentation and justification (HABERMAS, 1996, p. 4-5)[5].

Habermas acknowledges that everyday practices show that communicative actors are not always motivated by good reasons. Nevertheless, from this empirical standpoint, he maintains that the accountability of agents is—much like Kant's idea of freedom—a counterfactual presupposition. Conversely, "the supposition of rationality is a *defeasible* assumption and not *a priori* knowledge. It 'functions' as a multiply corroborated pragmatic presupposition that is constitutive of communicative action. But in any given instance, it can be falsified." (HABERMAS, 2003, p. 97) This implies that the assumption of rationality in communicative action "is open to being contradicted by experiences that participants have precisely through engaging in this practice." (HABERMAS, 2003, p. 99). Precisely because rationality constitutes a defeasible assumption rather than a priori knowledge, "discursive rationality owes its special position not to its foundational but to its integrative role." (HABERMAS, 1998, p. 309).

[5] On this topic, see also LAFONT, 1999, pp. 317.

This leads to the second essential caveat: none of the three structures of rationality is pre-linguistic in nature. Even epistemic and teleological rationality—despite their potentially monological or non-communicative forms—cannot be detached from linguistic mediation. Ultimately, both are linked to communicative rationality through the integrative function of discursive rationality. This conceptual link is clarified in the following terms:

> the reflection of the rational person who distances himself from himself, the rationality inherent in the structure and in the procedure of argumentation is *mirrored* in a general way. However, it becomes clear at the same time that on the integrative level of reflection and discourse, the three rationality components – knowing, acting and speaking – combine, that is, form a syndrome (HABERMAS, 1998, p. 311).

This multifaceted 'syndrome' underscores that rationality is not a monolithic entity, but a complex integration of functions. To better grasp how this integrative role of discursive rationality and the social nature of rationality operate in practice, the distinction between the communicative and non-communicative uses of language is particularly illuminating. In this view, non-communicative language is characteristic of epistemic and teleological rationalities; in these realms, linguistic usage does not depend on an interpersonal relationship between speaker and hearer within a communicative context. That is to say, language users are not pursuing illocutionary goals. Thus, the non-communicative use of language "for purposes of pure representation or for a plan of action played through mentally is due to a feat of abstraction that merely suspends the reference—which is *always present virtually*—of propositions to truth, or of intentions to the seriousness of what is resolved" (HABERMAS, 1998, p. 319).

Epistemic and teleological rationality may, therefore, employ language in both communicative and non-communicative modes. In contrast, communicative rationality is inherently dependent on the communicative use of language. Manifested through arguments and imperatives—rather than declarative or intentional propositions—this communicative use is pragmatic in origin, relying on interpersonal relations and illocutionary meaning. As Habermas clarifies, propositions and intentions "can be divested of the illocutionary meaning of acts of asserting and announcing without losing their meaning, whereas even in *foro interno* an imperative without an illocutionary component would no longer be an imperative." Accordingly, communicative rationality "is first embodied only in a process of reaching understanding that operates by way of validity claims whenever speaker and hearer, in a performative attitude directed to second persons, (want to) reach understanding with one another about something in the world" (HABERMAS, 1998, pp. 319-320).

In terms of the communicative use of language, a fundamental distinction is drawn between 'weak' and 'strong' modes of reaching understanding. This conceptual difference is clarified as follows:

> Now, of course, it makes a difference whether agreement (*Einverständnis*) concerning a fact exists between participants or whether they both merely reach an understanding (*sich verständigen*) with one another concerning the seriousness of the speaker's intention. *Agreement* in the strict sense is achieved only if the participants are able to accept a validity claim for the *same* reasons, while *mutual understanding* (*Verständigung*) can also come about when one participant sees that the other, in light of her preferences, has good reasons in the given circumstances for her declared intention—that is, reasons that are good *for her*—without having to make these reasons his own in light of his preferences (HABERMAS, 1998, p. 320-321).

The *weak mode of reaching understanding* is analyzed through declarations of intent and simple imperatives. Within this line of reasoning, statements such as "I will travel tomorrow" or commands like "sit down" do not seek to produce consensus, despite being illocutionary acts. This constitutes a weak sense of reaching understanding because, while validity claims are raised

and may be accepted or rejected, they do not require shared normative agreement. In the case of the declaration "I will travel tomorrow," for instance, the speaker gains assent by showing that the action is rational in light of her preferences; thus, teleological rationality assumes a mediating role. It is therefore sufficient for the hearer to have good reasons to trust the speaker's intent, even without sharing the underlying reasons.

Here, a distinction is drawn between "publicly intelligible reasons" (characteristic of the weak mode) and "generally acceptable reasons" (characteristic of the strong mode). In this way, intentional declarations and imperatives—notwithstanding their success-oriented nature—still "move within the horizon of a mutual understanding based on validity claims and thus still within the domain of communicative rationality." Their illocutionary success does not require raising claims to rightness; instead, it is "in turn measured in terms of claims to truth and truthfulness even if this is only with reference to the preferences of the speaker (…). The hearer assumes that the speaker means what she says and holds it to be true" (HABERMAS, 1998, p. 323).

The *strong mode of reaching understanding*, on the other hand, is explored through the analysis of promises, declaratives, and commands—speech acts that invoke normative validity claims. This mode emerges when the truth of statements is thematized or when they are situated within normative contexts that invite such thematization. For instance, the assertion "I will sign a contract tomorrow" may be understood merely as an intentional declaration; however, depending on the context or the line of questioning, it could signify a promise through which a speaker commits herself to an action. This shift in meaning is articulated as follows:

> the illocutionary meaning and validity basis of the utterances change. Normative reasons do not determine the prudential assessments of *arbitrary closing* decisionmaking subjects; they determine rather the decisions of subjects who *bind their wills* and are thus able to enter into obligations. In contrast to the case of 'naked' declarations of intentions and 'simple' imperatives, normative reasons are not actor-relative reasons for one's own (or another's) purposive-rational behavior—as in the case of assertions—actor-independent reasons; however, unlike the reasons for assertions, they are not reasons for the existence of states of affairs but rather for the satisfaction of normative binding expectations (HABERMAS, 1998, p. 324-325).

In this sense, promises, declaratives, and commands constitute speech acts that carry a claim to validity grounded in practical discourse. Grasping the illocutionary meaning of such acts requires familiarity with their specific normative context. This relationship is further clarified in the following terms: "[i]nsofar as the participants intersubjectively recognize a normative background (…), they can accept regulative speech acts as valid for the same reasons" (HABERMAS, 1998, p. 325).

It becomes clear, therefore, that not every use of language is communicative, nor does all linguistic communication seek to reach understanding in the strong sense. However, the most important element for conceptualizing social rationality—and its irrational counterparts—is that the use of language to reach understanding extends beyond the boundaries of communicative rationality to encompass epistemic and teleological rationalities as well. The core structures of rationality are thus inextricably linked to discursive praxis, as they consistently refer back to the level of argumentation: the sphere in which they are critically tested. This conceptual synthesis is captured in the following terms: "'practical reason' is not an elementary phenomenon but rather goes back to an entwinement—effected within the framework of social interactions—of epistemic and teleological rationality with communicative rationality" (HABERMAS, 1998, p. 325).

Thus, while the social character is most evident within the realm of communicative rationality, the mediating function of discursive rationality reveals how these structures are intertwined; consequently, any failure of epistemic or teleological justification may also be regarded as a manifestation of social irrationality. This sociocultural dimension of rationality is further articulated in the following terms:

> (…) the operations of 'reason' take the form of a circulation of linguistically linked reasons. For reasons circulate, so to speak, between their forms, which are symbolically consolidated in sociocultural ways of life, on the one hand, and the flow of communication and corresponding thought and consciousness processes of the acting subjects, on the other (HABERMAS, 2024, pp. 161-162).

In essence, for Habermas, rationality—across all its structures or roots—is inherently social by virtue of its anchorage in language. Consequently, irrationality, understood as the denial of the justifications required by each specific form of rationality, is likewise a social phenomenon.

II – Rationality, irrationality and justification

A question that arises is how effectively this concept of rationality can address the social, political, and legal problems caused by irrational practices. Discussing the connection between this theoretical framework and practical concerns in a recent series of interviews, Habermas asserts that: "it is precisely in questions of democratic and legal theory that the genuine power of practical reason proves its worth." (HABERMAS, 2024, p. 178-179). In the same vein, when addressing reason's potential to tackle social and political challenges in earlier political essays, specific phenomena are identified as inherently irrational. These include poverty, the threats posed by the arms race, the aggressive depletion of natural resources, and ecological instability, as well as the deprivation of rights and the vulnerabilities faced by individuals and minorities. In this context, he characterizes the task of reason in the following terms: "[r]eason is there to give voice to this negativity, to lend our voice to those who are silenced by pain, to 'bring the unreasonable to reason'—in opposition to the existing unreasonable (…)." (HABERMAS, 1990a, p. 84).

To what extent, however, can this concept of rationality generate obligations, and can it establish a normative basis for social, political, and legal practices? The answer is affirmative, though it entails a procedural requirement rather than a substantive one (cf. COOKE, 1994, p. 43). When applied to the spheres of morality, politics, and law, this procedural approach provides a "framework of acceptability" that functions as a filter for social irrationality. Within this framework, irrational content is identified and potentially excluded from norms of action insofar as agents fail to provide the justification required by the corresponding validity claims[6].

In what follows, I intend to explore the moral, political, and legal dimensions of social irrationality by focusing on the distinction between justifications connected to claims to truth and those pertaining to rightness.

As discussed above, rationality is treated as a presupposition that participants in argumentative processes must necessarily assume. Following the detranscendentalization of reason initiated by the linguistic turn, the criterion for the objectivity of knowledge is situated within the public justificatory praxis of a communication community (cf. HABERMAS, 2003, p. 249). Nevertheless, to avoid the pitfalls of contextualism and historicism, the aim is to show that rationality

[6] In previous research, I have explored how this framework is mobilized to address phenomena categorized as social irrationality—most notably hate speech. See CONSANI, 2025, p. 191; CONSANI, 2023, p. 562; for a contrasting perspective on this topic, see also GALUPPO; SILVA, 2021, p. 131.

resides within a perspective of immanent- transcendence. According to this view, "participants in communication can neither understand nor misunderstand one another unless there is a presupposition of rationality" (HABERMAS, 2003, p. 86; p. 93). It is precisely upon these presuppositions that the possibility of transcendence is grounded. In *Truth and Justification*, at least four presuppositions essential to establishing the relationship between immanence and transcendence are outlined: (i) the common objective world; (ii) the accountability of acting subjects; (iii) the unconditionality of validity claims; and (iv) discourse as the ultimate forum of justification.

(i) *The common objective world*: According to Habermas, "[d]etranscendentalization leads, on the one hand, to the embedding of knowing subjects into the socializing context of a lifeworld and, on the other hand, to the entwinement of cognition with speech and action" (HABERMAS, 2003, p. 88-89). This implies that to reach an understanding about something in the world, subjects must necessarily operate from within their shared lifeworld. Yet, in doing so, they presuppose "'the world' as the totality of independently existing objects that can be judged or dealt with." The objective world is thus defined as one that 'is "given" to us as "the same for everyone"' (HABERMAS, 2003, p. 89).

The idealist transcendental perspective is thus superseded by a commitment to internal realism. While the former "conceives the totality of objects of possible experience as a world 'for us,' as a world of appearances," the latter acknowledges that "everything that can be represented in true statements is 'real,' although facts are interpreted in a language that is always 'ours'" (HABERMAS, 2003, p. 90). In this sense, the orientation toward truth acquires an essential regulative function for fallible justificatory processes. Drawing a comparison to Kantian philosophy, the following clarification is provided:

> Even after objective knowledge is detranscendentalized and tied to discursive justification as the 'touchstone of truth,' the point of Kant's injunction against the apodictic use of reason and the transcendent use of the understanding is preserved. Only now the boundary separating the transcendental from the transcendent use of our cognitive capacity is not defined by sensibility and understanding, but by the forum of rational discourse in which the convincing power of good reasons must flourish (HABERMAS, 2003, p. 92).

The orientation toward truth thus serves a regulative function, grounded in both the presupposition of a shared objective world and the requirement for justifications rooted in the lifeworld. Drawing further parallels with Kantian thought, Habermas maintains that "in the course of detranscendentalization, the theoretical ideas of reason step out of the static 'intelligible world' and unleash their dynamics within the lifeworld." (HABERMAS, 2003, p. 92). Immanent-transcendence implies that disputes over the correct interpretation of the world must be transcended from within—that is, from the perspective of participants who interpret their world from the situated horizon of their habits and traditions.

(ii) *The accountability of acting subjects*: This presupposition was introduced previously during the analysis of the scope of communicative rationality. Habermas emphasizes that while the presupposition of a shared objective world addresses the relationship between subjects and objects from a descriptive perspective, the accountability of acting subjects opens onto the normative dimension of the social world. This entails a reciprocal attribution of rationality that agents must grant one another when engaging in communicative action. Participants assume that their interlocutors are both rational and accountable for their utterances and actions. As such, they are expected to justify their positions by mobilizing rationality across its epistemic, teleological, and

communicative structures—drawing upon the full spectrum of validity claims (truth, truthfulness, and rightness) within an argumentative process that remains inherently open to fallibility. (cf. HABERMAS, 2003, p. 93-99)

(iii) *The unconditionality of validity claims*: Within the scope of formal pragmatics, transcendental projection carries a weak, detranscendentalized sense. This perspective distinguishes between a vertical dimension of world-relation—where idealization consists of anticipating the totality of possible references—and a horizontal dimension of intersubjective relations, characterized by the mutual assumption of rationality between subjects. Validity claims are understood in this same light, under the premise that "[i]f reaching understanding, and thereby coordinating action, is to be possible at all, then agents must be capable of taking a warranted stance on criticizable validity claims and of orienting themselves by such claims in their own actions" (HABERMAS, 2003, p. 99).

Regarding validity claims, idealization involves a preliminary abstraction from deviations, individual differences, and limiting contexts. Here, the tension between immanence and transcendence reveals a dual movement. On the one hand, "[t]he presupposed objectivity of the world is so deeply entwined with the intersubjectivity of reaching an understanding about something in the world that we cannot transcend this connection and escape the linguistically disclosed horizon of our intersubjectively shared lifeworld." Conversely, this situatedness does not preclude universal reach, as "[w]e are able reflectively to transcend whatever our given initial hermeneutic situations are and attain intersubjectively shared views on disputed matters" (HABERMAS, 2003, p. 100). Although the justificatory processes for truth and rightness differ—as examined below—the discursive process remains the vital link, given that it "increases the responsive potential by which rationally accepted claims to validity prove their worth." (HABERMAS, 2003, p. 102)

(iv) *Discourse as the ultimate forum of justification:* In analyzing this assumption, Habermas demonstrates how the previous presuppositions converge. In particular, he highlights the relationship between validity claims in the objective and social worlds, noting that it is within the forum of discourse that agents refer to objects through the propositional content of their statements while simultaneously addressing norms as elements of the social world. In this context, he defines rational discourse as "a process that ensures the inclusion of all those affected and the equal consideration of all the interests at play." Such equality plays a crucial role in the pursuit of understanding because "in view of the idea that only those norms equally good for all merit recognition from the moral point of view, such discourse presents itself as the appropriate method of conflict resolution" (HABERMAS, 2003, p. 105).

To ensure that the discussion of contested validity claims within a discourse does not lose its cognitive purpose, participants must subscribe to a structurally mandatory egalitarian universalism. At first glance, this universalism carries a formal-pragmatic meaning rather than a moral one. In this context, the rational acceptability of validity claims is ultimately grounded in reasons capable of withstanding objections under demanding communicative conditions. Habermas recognizes that "if this is the intuitive meaning that we associate with argumentation in general, then we also know that a practice may not seriously count as argumentation unless it meets certain pragmatic presuppositions" (HABERMAS, 2003, p. 106).

There are four unavoidable pragmatic presuppositions at play, namely: *inclusivity* (whereby

those capable of making relevant contributions cannot be excluded); *the equal distribution of communicative freedoms* (granting equal opportunity to contribute); *truthfulness* (the requirement that participants express their sincere thoughts); and *the absence of external or internal constraints* (ensuring that positions are motivated solely by the force of the arguments themselves) (cf. HABERMAS, 2008, p. 82; 2003, p. 106-107). Admittedly, "[t]hese argumentative presuppositions obviously contain such strong idealizations that they raise the suspicion of a rather tendentious description of argumentation;" nevertheless, it is essential to recognize that these presuppositions, "no matter how counterfactual, are by no means mere constructs. Rather they are operatively effective in the behavior of the participants themselves. Someone who seriously takes part in an argument de facto proceeds from such presuppositions" (HABERMAS, 2003, p. 107-108).

Accordingly, the aforementioned presuppositions function within this theoretical scope as a 'framework of acceptability' connected to the very structure of language. Although these assumptions appear highly idealized, the reconstruction of rationality within social practices reveals their presence whenever one engages in argumentation aimed at mutual understanding—even if only counterfactually. This matter is taken up below through an analysis of the distinction between truth and justification, as well as the differentiation between justifications concerning claims to truth and those concerning claims to rightness.

Regarding the relationship between truth and justification, Habermas's theory has changed over the years, shifting from an epistemic and anti-realist perspective to a non-epistemic view with realist elements. An epistemic conception holds that the truth of a proposition depends on its justification. Such perspectives are commonly associated with anti-realism, since propositional truth is not seen as depending on things being as the propositions state they are. One problem with these anti-realist stances is that they may lead to contextualism or relativism. In this sense, it becomes more difficult to address issues of social irrationality, as evaluative criteria can be rendered more flexible by the justifications presented. A non-epistemic perspective, in turn, maintains that the truth of a proposition does not depend on whether someone has a justification for believing it. This view is generally associated with realism, as it is consistent with the understanding that propositional truth depends not on us, but on the world (cf. ZUIDERVAART, 2017, p. 103).

Zuidervaart identifies three distinct stages in the development of Habermas's concept of truth: a) *consensus theory* (developed in the early 1970s, wherein Habermas proposed an epistemic conception of truth, explaining it through the conditions under which truth claims are justified); b) *formal pragmatics* (beginning with *The Theory of Communicative Action*, Habermas distinguishes more clearly between truth and justification. In this stage, he replaces his earlier consensus theory with a formal-pragmatic theory of meaning rather than focusing on truth per se. Thus, "in place of a consensus theory of truth, he proposes a formal pragmatic theory of meaning"); and c) *pragmatic realism* (since the 1996 publication of his essay *Richard Rorty's Pragmatic Turn*, Habermas has emphasized the distinction between truth and justification, as well as their internal relationship, arguing that truth cannot be reduced to rational assertibility. In *Truth and Justification*, he further develops his defense of the 'Janus-faced' nature of truth—incorporating realist elements concerning truth and anti-realist elements regarding justifica-

tion) (cf. ZUIDERVAART, 2017, p. 105)[7].

In Truth and Justification, Habermas himself identifies two distinct concepts of truth within his work: the first is a procedural definition—the discursive concept of truth; later, he moves toward defending a pragmatic concept. The discursive concept favored identifying truth with ideal or rational assertibility[8] as a way to escape the dilemmas that arise when one recognizes, through formal pragmatics, that the reality we encounter is not 'naked,' but already permeated by language. Within this context, the challenge lies in upholding claims to universal truths that transcend context without reverting to a realist perspective that denies that our knowledge of truth is mediated by linguistic and social interactions. This tension leads to the following observation:

> The attempt to combine the language-transcendent understanding of reference with a language-immanent understanding of truth as ideal assertibility promised a way out of this dilemma. On this view, a statement is true if and only if, under the rigorous pragmatic presuppositions of rational discourse, it is able to withstand all efforts to invalidate it, that is, if and only if it can be justified in an ideal epistemic situation. Inspired by C. S. Peirce's famous suggestion, K.-O. Apel, H. Putnam, and I have all at one time or another defended some version of such *a discursive concept of truth* (HABERMAS, 2003, p. 36, italics added).

From this perspective, the meaning of truth was tested within argumentative praxis by appealing to the unavoidable pragmatic presuppositions of discourse—namely, inclusivity, the equal distribution of communicative freedoms, truthfulness, and the absence of external or internal constraints. Thus, on the one hand, the discursive concept of truth was developed "to take account of the fact that a statement's truth—absent the possibility of direct access to uninterpreted truth conditions—cannot be assessed in terms of 'decisive evidence,' but only in terms of justificatory, albeit never definitively 'compelling,' reasons." On the other hand, to avoid falling into contextualism, it was maintained that an "idealization of certain features of the form and process of the practice of argumentation was to characterize a procedure that would do justice to the context-transcendence of the truth claim raised by a speaker in a statement by rationally taking into account all relevant voices, topics, and contributions" (HABERMAS, 2003, p. 37).

In reassessing his theory, Habermas considers the discursive concept of truth not as incorrect, but as insufficient, since it fails to explain what authorizes us to regard a supposedly ideally justified statement as true (cf. HABERMAS, 2003, p. 252). Furthermore, he argues that the discursive conception is counterintuitive, insofar as truth is not a concept linked to success; instead, a proposition "is agreed to by all rational subjects because it is true; it is not true because it could be the content of a consensus attained under ideal conditions" (HABERMAS, 2003, p. 101). Given these challenges, in dialogue with critics such as Wellmer (1992) and Lafont (1999), he acknowledges a gap between truth and rational assertibility, stemming either from the inherent fallibility of justification itself or from the fact that the conditions necessary to eliminate such fallibility remain beyond human reach (cf. HABERMAS, 2003, p. 38). As he notes: "[t]hese objections have prompted me to revise the discursive conception of rational acceptability by

[7] See also FULTNER, 2019, p. 446-449; STRYDOM, 2019, p. 555.

[8] In his analysis of the differentiation of rightness from truth in Habermas's work, Strydom emphasizes the distinction between the process of argumentation (assertibility) and its result or success (acceptability). *Rational assertibility* acts as a regulatory idea that guides the direction and conduct of the argumentative process, referring to the effort to defend a proposition in accordance with the internal requirements of the matter in question. *Rational acceptability*, on the other hand, concerns the closure or conclusion of the process, representing the "achievement" of the argument—that is, the moment when the claim of validity is effectively recognized and accepted by the participants after discursive scrutiny. Despite their differences, both concepts can be characterized as forms of ideal justification. (cf. STRYDOM, 2019, p. 560-562)

relating it to a *pragmatically conceived, nonepistemic concept of truth*, but without thereby assimilating 'truth' to 'ideal assertibility'" (HABERMAS, 2003, p. 38, italics added).

The pragmatic concept of truth, in turn, establishes a distinction between truth and rational assertibility or justification. Habermas continues to assert that there is a connection between truth and justification that is "*epistemically* necessary (i.e., we can only come to know what is true by means of providing reasons)" but "not *conceptually* necessary (i.e., truth cannot be defined in terms of justification or vice versa)" (FULTNER, 2019, p. 447; cf. HABERMAS, 2003, p. 38). Thus, although the gap between truth and rational assertibility cannot be bridged theoretically within discourse, it is bridged pragmatically through action (cf. HABERMAS, 2003, p. 92). Since participants in an interaction cannot suspend their truth claims, these claims function as certainties that guide their actions (cf. HABERMAS, 2003, p. 252-253). In this way, "the pragmatic role of a Janus-faced truth that establishes the desired internal connection between performative certainty and warranted assertibility" is revealed (HABERMAS, 2003, p. 253).

From this perspective, the objectivity of knowledge is ensured because the objective world links truth to reference, connecting the truth of statements with the objectivity of what is stated. The objective world guarantees this objectivity either by imposing a limit—a certain unavailability—on the range of possible interpretations, or by being the same for everyone. Today, for instance, one can no longer truthfully claim that the Earth is flat; such a claim is no longer open to interpretation, as we are all referring to the same objective world. These elements of unavailability (*Unverfügbarkeit*) and identity (*Identität*) are highlighted in the following passage:

> The concept of the 'objective world' encompasses everything that subjects capable of speech and action do not 'make themselves' irrespective of their interventions and inventions. This enables them to refer to things that can be identified as the same under different descriptions. The experience of 'coping' accounts for two determinations of 'objectivity': the fact that the way the world is not up to us; and the fact that it is the *same* for all of us. Beliefs are confirmed in action by something different than in discourse (HABERMAS, 2003, p. 254)[9].

Thus, in the realm of action, the unavailability and identity of the objective world safeguard the objectivity of justifications, as these are tested through their contact with the world. This demarcation of objectivity is what connects the pragmatic conception of truth with realist elements. In short, this perspective is pragmatic because it maintains a link between truth and justification, recognizing that we only access truth through the provision of reasons. At the same time, it incorporates realistic elements by anchoring the concept of truth in the unavailability and identity of the objective world. These elements prevent truth from being reduced to mere justification, as the objective world itself offers resistance and puts our claims to the test.

In the realm of discourse, however, the process functions somewhat differently. The varying roles of justification within the spheres of action and discourse are better understood by analyzing the distinction between claims to truth and claims to rightness. This distinction is addressed in the following passage:

> Moral validity claims lack the reference to the objective world that is characteristic of claims to truth. This means they are robbed of a justification-transcendent point of reference. The reference to the world is replaced by an orientation toward extending the borders of the social community and its consensus about values. If we want to specify the difference between rightness and truth more precisely, we have to examine

[9] There is a stronger emphasis on the concepts of "unavailability" (*Unverfügbarkeit*) and "identity" (*Identität*) in the original German than in the English translation. Cf: "Unverfügbarkeit und Identität ler Welt sind die beiden Bestimmungen von 'Objektivität', die sich us der Erfahrung des 'Coping' erklären: Überzeugungen 'bewäh-em' sich im Handeln an etwas anderem als im Diskurs" (HABERMAS, 2004, p. 321).

> whether and, if so, how this orientation toward an ever more extensive inclusion of other claims and persons can make up for the missing reference to the world (HABERMAS, 2003, p. 257).

The consensus reached through discourse carries different implications for the truth of empirical statements than it does for the rightness of moral judgments and norms. In the context of the objective world, the truth of a statement simultaneously denotes a fact. Facts owe their factuality to being rooted in a world of objects that exist independently of any description. This interpretation implies that a consensus on a statement may prove false in light of new evidence, no matter how carefully reached or well-founded it may be (the flat-earth thesis again serving as an example). As noted above, within the realm of truth claims, objectivity is ensured by an objective world circumscribed by the conditions of unavailability and identity.

Regarding claims to rightness, the distinction between truth and ideal warranted assertibility disappears. In the case of moral validity, there is no equivalent to the ontological interpretation of validity that characterizes truth. To highlight this difference, Habermas points out that "[w]hereas successful learning in the sphere of empirical problems may *result* in agreement, learning in the moral domain is *assessed* in terms of how inclusive such a consensus reached through reason-giving is." (HABERMAS, 2003, p. 257). Unlike claims to truth, the consensus reached through discourse does not establish facts; rather, it establishes a norm that must merit intersubjective recognition. Those involved proceed from the assumption that such recognition can be secured under the approximately ideal conditions of rational discourse. Consequently, the validity of a normative statement is not understood "in terms of the *obtaining* of a state of affairs, but as the *worthiness of recognition* of a corresponding norm on which we ought to base our practice. A norm worthy of being recognized cannot be denied by a 'world' refusing to 'play along.'" (HABERMAS, 2003, p. 257-258). Thus, while truth remains a non-epistemic concept—established within the context of justificatory practices but not reducible to their results—rightness functions differently: "[s]ince the 'validity' of a norm consists in that it would be accepted, that is, recognized as valid, under ideal conditions of justification, 'rightness' is an epistemic concept" (HABERMAS, 2003, p. 258).

If rightness is an epistemic concept, what ensures the unconditionality and universality of claims to rightness? Regarding unconditionality, this is grounded in the criterion of inclusivity. While the objective world is circumscribed by the determinations of unavailability and identity, the social world is bound solely by the determination of identity; this, in turn, necessitates the equal inclusion of all claims and individuals. In the social dimension, participants must construct an inclusive 'we-perspective' and promote the reciprocal adoption of perspectives. According to Habermas,

> Following this constructivist conception, the unconditional nature of moral validity claims can be accounted for in terms of the universality of a normative domain that *is to be brought about*: Only those judgments and norms are valid that could be accepted for good reasons by everyone affected from the inclusive perspective of equally taking into consideration the evident claims of all persons (HABERMAS, 2003, p. 261).

Universality, in turn, is also ensured by the egalitarian nature of rules that serve the equal interests of all those affected. This is supplemented by the perspective of procedural justice, which requires that rules be justified and applied impartially (cf. HABERMAS, 2003, p. 264).

III - Final Remarks

This article examines the concept of social irrationality in Habermas's work to assess its po-

tential for addressing contemporary forms of irrationality. It begins by outlining his account of rationality to clarify what may properly be regarded as irrational. Given that Habermasian concepts are rooted in formal pragmatics—which analyzes the rationality embedded in language and social practices—rationality across all its structures (epistemic, teleological, and communicative) is intrinsically social. However, it is in communicative rationality that this social character manifests most prominently. At its core, rationality is defined by the requirement to provide justifications for the validity claims raised in argumentation, whether these concern knowledge, action, or speech. Irrationality, accordingly, arises when an interlocutor fails to provide the expected justification for a given claim. Habermas thus associates irrationality with dogmatism or the absence of justification.

This concept of rationality is operationalized through specific idealizations—namely, presuppositions regarding our relationship to the objective and social worlds. These pragmatic assumptions underpin a "framework of acceptability" in which justifications are offered and validity claims determine what is deemed rational. Thus, Habermas's distinction between claims to truth and claims to rightness is crucial for identifying and addressing phenomena of social irrationality.

To conclude, I will test these concepts against specific instances of social irrationality. First, one can examine truth claims within the context of scientific denialism during the COVID-19 pandemic. At the onset of the crisis, investigating the efficacy of existing drugs, such as hydroxychloroquine, was a legitimate scientific hypothesis and thus not irrational. However, the persistent advocacy for such treatments—even after robust clinical evidence had proven their ineffectiveness—marked a clear departure from epistemic rationality. Figures such as the French physician Didier Raoult, who became a global proponent of unproven therapies, failed to provide the justifications required for their claims to truth.

Drawing on Habermas's theory of rationality, this behavior can be considered social irrationality because it violates the presupposition of an identical and unavailable objective world. By disregarding the "resistance" offered by empirical data, these actors retreated into dogmatism. When a validity claim to truth is raised despite a clear lack of grounding in empirical facts, it ceases to be an invitation to discourse and becomes an ideological imposition. In this sense, the irrationality of the chloroquine defense lies not in the initial hypothesis but in the refusal to abandon a claim that the objective world had already proven false. In this case, dogmatism supplanted the rational requirement for justification.

In the socio-political sphere, Habermas identifies—as noted earlier—specific irrational phenomena, such as poverty and the deprivation of rights. These conditions fail to find justification on the basis of claims to rightness. According to this framework, the primary criterion for validating a norm is its "worthiness of recognition," a status that is inextricably linked to the equal inclusion of all persons and their claims. Applying these criteria makes it clear that social phenomena such as poverty or the deprivation of rights cannot be validated within the scope of moral or legal norms. Furthermore, these criteria are valuable in assessing the legitimacy of social struggles, allowing for a distinction between progressive movements that strive for inclusion and regressive[10] groups that seek to maintain privilege and discrimination.

[10] For a discussion on democratic regression in Habermas's theory, see WOLKENSTEIN, 2025.

Building on this distinction, one can analyze the disparate legitimacy of movements such as Black Lives Matter, the January 6th Capitol riot, and the January 8th Brasília attacks through the lens of the claims to rightness. The Black Lives Matter movement is grounded in a struggle for universal inclusion and the recognition of rights systematically denied to a marginalized group; its goals align with the "worthiness of recognition" as they seek to expand the democratic "we-perspective." In contrast, the invasion of the Capitol and the attacks in Brasília represent a regressive effort to undermine democratic institutions and exclude the voices of a legitimate majority in favor of maintaining a particularistic privilege—the attempt by a specific group to assert its own will, identity, or political preference as superior to the universal rules of the democratic process.

A greater challenge arises when social movements seeking inclusion and recognition also resort to violence, public disorder, or property damage—as occurred in isolated instances during the Black Lives Matter protests. In such cases, even if the underlying claim is initially legitimate, the use of force represents a rupture with the communicative pursuit of mutual understanding, thereby undermining its legitimacy. Nevertheless, Habermas's theory of rationality provides the normative benchmarks necessary to distinguish legitimate social struggles from regressive ones. Furthermore, it offers a robust framework for identifying when the actions of a progressive movement transgress the discursive boundaries essential to democratic dialogue.

Bibliographic References

BENHABIB, S. 1986. *Critique, norm and utopia*. Nova York: Columbia University Press.

CELIKATES, R. 2018. *Critique as Social Practice*: Critical Theory and Social Self-Understanding. Trans. Naomi van Steenbergen. New York/London: Rowman & Littlefield International.

CONSANI, C. F. 2023. Liberdade de expressão, discursos de ódio e irracionalidade social: uma análise a partir da teoria da democracia habermasiana. *Ethic@*, v. 22, n. 2, p. 562-596.

CONSANI, C. F. 2025. Freedom of expression and hate speech in Habermas's democratic theory. In: CONSANI, C. F.; DUTRA, D. V.; KLEIN, J. (Eds). *Concepts and Conceptions of Freedom*: Historical perspectives and contemporary developments. 1. ed. Münster: Lit Verlag.

COOKE, M. 1994. *Language and Reason*. Cambridge, MA: MIT Press.

DUTRA, D. V. 2005. *Razão e Consenso em Habermas*. Florianópolis: Editora da UFSC.

FINLAYSON, J. G. 2013. The Persistence of Normative Questions in Habermas's Theory of Communicative Action. *Constellations*, v. 20, n. 4, p. 518-532.

FULTNER, B. 2019. Truth. In: ALLEN, A.; MENDIETA, E. (Eds). *The Cambridge Habermas Lexicon*. Cambridge and New York: Cambridge University Press.

GALUPPO, M. C.; SILVA, B. B. 2021.Tolerância, liberdade de expressão e a esfera pública em Habermas. *Dois pontos*, Curitiba, São Carlos, v. 18, n. 2, p. 131-145.

HABERMAS, J. 1976. *Communication and the Evolution of Society*. Translated by Thomas McCarthy. Boston: Beacon.

HABERMAS, J. 1984. *The Theory of Communicative Action*: reason and the rationalization of society. Volume I. Translated by Thomas McCarthy. Boston: Beacon Press.

HABERMAS, J. 1990. *The Philosophical Discourse of Modernity*. Translated by Frederick Lawrence. Cambridge: Polity Press.

HABERMAS, J. 1990a. *Die Nachholende Revolution*. Berlin: Suhrkamp Verlag, 1.

HABERMAS, J. 1996. *Between Facts and Norms*: Contributions to a Discourse Theory of Law and Democracy. Translated by William Rehg. Cambridge: The MIT Press.

HABERMAS, J. 1998. *On the pragmatics of communication*. Cambridge: The MIT Press.

HABERMAS, J. 2003. *Truth and Justification*. Translated by Barbara Fultner. Cambridge: The MIT Press.

HABERMAS, J. 2004. *Wahrheit und Rechtfertigung*. Frankfurt am Main: Suhrkamp Verlag.

HABERMAS, J. 2008. *Between Naturalism and Religion:* Philosophical Essays. Translated by Ciaran Cronin. Cambridge: Polity Press.

HABERMAS, J. 2024. *Es musste etwas besser werden*. Gespräche mit Stefan Müller-Doohm und Roman Yos. Berlin: Suhrkamp Verlag.

ISER, M. Rationale Rekonstruktion. 2009. In: BRUNKHORST, H.; KREIDE, R.; LAFONT (Eds.). *Habermas-Handbuch*. Stuttgart: Metzler.

LAFONT, C. 1999. *The Linguistic Turn in Hermeneutic Philosophy*. Cambridge, MA: MIT Press.

LAFONT, C. 2009. Communicative Rationality. In: BRUNKHORST et al (Eds.). *The Habermas Handbook*. New York: Columbia University Press.

MCCARTHY, T. 1994. Kantian Constructivism and Reconstructivism: Rawls and Habermas in Dialogue. *Ethics*, v. 105, n. 1, p. 44-63.

NOBRE, M.; REPA, L. 2012. Reconstruindo Habermas: etapas e sentido de um percurso. In: NOBRE, M.; REPA, L. (Eds.). *Habermas e a reconstrução*: sobre a categoria central da Teoria Crítica habermasiana. Campinas: Papirus.

PEDERSEN, J. 2008. Habermas' method: rational reconstruction. *Philosophy of social sciences*, v. 38, n. 4, p. 457-485, dez 2008.

PETERS, B. 1994. On reconstructive legal and political theory. *Philosophy & Social Criticism*, v. 20, n. 4, p. 101-134.

REED, I.; MOORE, A. C. 2019. Rationality/Rationalization. In: ALLEN, A.; MENDIETA, E. (Eds). *The Cambridge Habermas Lexicon*. Cambridge and New York: Cambridge University Press.

REPA, L. 2008a. *A transformação da filosofia em Jürgen Habermas*: os papéis de reconstrução, interpretação e crítica. São Paulo: Singular.

REPA, L. 2008b. Jürgen Habermas e o modelo reconstrutivo de teoria crítica". In: NOBRE, M. (Ed.). *Curso livre de Teoria Crítica*. Campinas: Papirus.

REPA, L. 2012. Reconstrução da história da teoria: observações sobre um procedimento da Teoria da ação comunicativa. In: NOBRE, M.; REPA, L. (Eds.). *Habermas e a reconstrução*: sobre a categoria central da teoria crítica habermasiana. Campinas: Papirus.

REPA, L. 2017. Compreensões de reconstrução: sobre a noção de crítica reconstrutiva em Habermas e Celikates. *Transformação*, v. 40, n. 3, p.10-28.

REPA, L. 2021. *Reconstrução e Emancipação*: Método e Política em Jürgen Habermas. São Paulo: UNESP.

SILVA, F. G.; MELO, R. 2012. Crítica e reconstrução em direito e democracia. In: NOBRE, M; REPA, L. (Eds). *Habermas e a Reconstrução*: sobre a categoria central da teoria crítica habermasiana. Campinas: Papirus.

STRECKER, D. 2019. Communicative Rationality. In: Allen, A; Medieta, E. (Eds.). *The Cambridge Habermas Lexicon*. Cambridge and New York: Cambridge University Press.

STRYDOM, P. 2019. On Habermas's differentiation of rightness from truth: Can an achievement concept do without a validity concept?. *Philosophy & Social Criticism* v.45, n. 5, p. 555–574.

WELLMER, A. 1992. What is a Pragmatic Theory of Meaning? In: HONNETH, A. et al., (Eds). *Philosophical Interventions in the Unfinished Project of Enlightenment*. Trans. William Rehg. Cambridge: MIT Press.

WOLKENSTEIN, F. 2025. The Dialectic of Backsliding: Thinking with Habermas About Democratic Progress and Regression. *European Journal of Philosophy*, v. 33, n. 4, p. 1274–1290.

ZUIDERVAART, L. 2017. *Truth in Husserl, Heidegger, and the Frankfurt School*. Cambridge, MA: MIT Press.

Rawls, temporal discontinuity and disasters

Rawls, descontinuidade temporal e desastres

Charles Feldhaus
Universidade Estadual de Londrina (UEL)
charlesfeldhaus@gmail.com

Abstract: This study aims to apply key concepts from Rawls's theory of justice as fairness to issues of justice in the context of disasters and catastrophes. Disasters can be understood as a form of temporary discontinuity, whereas catastrophes involve a prolonged disruption of people's normal lives. Whether an event becomes a disaster, a catastrophe, or neither, depends on the degree of preparedness in place—preparedness that can prevent what Kant, in *The Metaphysics of Morals*, refers to as cases of necessity. This study argues, based on aspects of Rawls's justice as fairness, that it is a requirement of justice for the state to adopt measures to mitigate temporal discontinuities and their harmful effects on the rational life plans of individuals understood as free and equal citizens.

Keywords: catastrophe; disaster; ethics; justice; temporal discontinuity.

Resumo: Este estudo pretende aplicar alguns conceitos da concepção da justiça como equidade de Rawls a questões de justiça em eventos de desastres e catástrofes. Desastres podem ser compreendidos como um tipo de descontinuidade temporária em oposição a uma catástrofe que consiste num tipo de descontinuidade prolongada da vida normal das pessoas. A diferença entre um evento se tornar um desastre ou uma catástrofe ou até mesmo nenhum dos dois depende de uma preparação adequada que impede que se caia naquilo que Kant denominou em *A metafísica dos costumes*, de casos de necessidade e esse estudo defende com base em aspectos da justiça como equidade de Rawls que é uma exigência de justiça que o Estado adote certas medidas contra a descontinuidade temporal e seus efeitos negativos aos planos racionais de vida de pessoas compreendidas como cidadãos livres e iguais.

Palavras-chave: catástrofe; desastre; ética; justiça; descontinuidade temporal.

Recebido em 24 de julho de 2025. Aceito em 11 de dezembro de 2025.

> "*Following a catastrophe, the effect of inequalities is everywhere and always the same: those who are vulnerable become more vulnerable, the poor are disenfranchised, while the rich become richer. Catastrophes entrench and augment inequalities that function to the detriment of those who are least advantaged and to the advantage of those who are already most advantaged.*" (Paul Dumouchel)

We are currently experiencing a particular feeling, a sense of insecurity about the future, and as a result, what we generally call our rational life plans (to use a now classic term from John Rawls' conception of justice) often seem to lose their meaning in the face of a scenario in which we are constantly bombarded by information that calls into question the future continuity of these plans. It is no coincidence that Byung-Chul Han wrote a book whose title is emblematic of the times we live in: *The Spirit of Hope*. In this book against the society of fear, he states that "fear circulates like a specter. We are frequently confronted with apocalyptic scenarios" (HAN, 2024, p. 9). In other words, until quite recently, apocalyptic scenarios were common only on movie screens, or would sometimes appear in newspaper reports showing that some people were experiencing disasters or catastrophes of greater or lesser magnitude. However, something seems to have changed in recent years, and it can be even be stated that something that was common only in the human imagination is now part of newspaper reports almost daily. Some might even say that this may simply be the result of the greater speed and near-simultaneity with which information, from the most remote places on the planet, is transmitted today, using new technologies, especially digital media. The problem is that new digital media are not only allowing immediate access to information about disasters or catastrophes around the globe, but through the dissemination of false or inaccurate information, what we normally call *fake news*, and through the algorithmic manipulation of people, they are also largely producing a feeling that we live in an era in which disasters and catastrophes are no longer the exception, but have become the rule. This has produced in people's minds a feeling of insecurity regarding the future of their rational life plans. In other words, the current scenario seems averse to the idea of temporal continuity, as we often experience the feeling that disasters and catastrophes are always imminent. The information we receive almost daily, through, mainly but not exclusively, new digital media, fosters the perception of the maintenance of temporal discontinuity. What makes this even worse is that those subject to this type of pessimistic information are already in an unfavorable situation regarding future planning. People in the most vulnerable conditions have less resilience in the face of disasters and catastrophes. This type of event hits the already disadvantaged hardest, those whom Rawls called the least favored in society in terms of justice as fairness.

However, the fact is that disasters and catastrophes often significantly affect the normal life of a society and it could be said that they call into question our ability to undertake our rational life projects, especially because they cause what we could call a situation of temporary or permanent temporal discontinuity. One could argue that sometimes we experience the imminence of a world war as a result of the actions of certain countries' governments; sometimes we experience the imminence of a climate crisis due to the behavior of governments and populations, which can jeopardize the survival of the human species on Earth; or sometimes we experience the imminent proliferation of a virus that could in a short space of time expand on a global scale, into a pandemic, which, even if it does not exterminate all people from the face of the earth, would negatively affect our lives and our loved ones and make long-term planning lose its meaning.

The philosophical debate about disasters

Ethical reflection on disasters and catastrophes can be undertaken based on different disciplines and areas of human knowledge, such as philosophy (as a matter of moral philosophy or ethics and political philosophy); theology (as the question regarding divine goodness in the face of the existence of evil in the world); the law (such as the question of responsibilities for the consequences of these types of events); the economy (such as the question of the cost-benefit calculation of investing resources in preparation and how this practice reduces the costs of response and reconstruction after the event, for example, on how these events have asymmetric effects on the more and less fortunate people in society); public health (such as the question of how to promote and care for people's health in a scenario that is often limited by scarce resources, that is, with the establishment of triage rules); sex studies (such as the question of the asymmetric effects on different sexes in a disaster event, when women often suffer these effects more than men), and so on. This type of interdisciplinary approach to disaster events and catastrophes can bring problems with regard to the definition of terms, since the way in which one of the disciplines or areas usually defines the terms may be different from the way in which another discipline defines the terms, and this may not always be without consequences with regard to dealing with the events, particularly considering preparation and response and the process of rebuilding the affected areas, which occurs after the event has passed and the immediate response is complete.

As Mathuna & Gordijn (2020, p. 2) argue, the debate regarding the definition of a disaster presents different views on the term, but ultimately it is not just an academic concern, since the use of different concepts and definitions of disasters (or catastrophes) have concrete impacts and many institutions that carry out humanitarian actions in response to this type of event are based on different concepts or definitions, which guide their conduct in the preparation, response, and recovery of affected locations after the event occurs. The problem is that there is little definitional and conceptual unity in legal documents and guidelines regarding disasters and catastrophes. However, in order to respond appropriately to these types of events, it is very important to at least outline the general features necessary for an event to be considered a disaster and a catastrophe or not. Here, we intend to show how the concept of temporal discontinuity, although unable to resolve the question, can introduce some aspects that are generally little considered in the debate about the definition.

In her book *Ethics for Disaster*, Naomi Zack defines a disaster as "an event (or series of events) that harms or kills a significant number of people or otherwise severely impairs or disrupts their daily lives in civil society. Disasters can be natural or the result of accidental or deliberate human action" (ZACK, 2009, p. 7). Furthermore, she adds that disasters cause surprise and shock, are unwanted by those they affect, cannot be said to be unpredictable, and also tend to generate media narratives about heroism and loss of life of those affected by disasters, etc. Naomi draws attention to the temporal discontinuity in people's normal lives and argues that inadequate preparedness can create a discontinuity between morality in normal times and morality or ethics in disasters. Good preparedness is a necessary, although not always sufficient, condition for mitigating the temporal discontinuity between normal life and life during a disaster event, and a necessary condition for preventing moral values (principles and virtues) that are valid in normal life from remaining valid in times of disaster. As it is only a necessary condition, it is important to emphasize that it is not always a sufficient condition, since in a disaster it is never impossible, even after the best

possible preparation, that unforeseen circumstances arise, and this inevitably makes it difficult to employ the morality of normal life in cases of disasters, when it may be necessary to apply, for example, triage practices to select who should live and who, because of the extremely adverse circumstances, may unfortunately have to be left to die or suffer more significantly the damage of the disasters to their health because of insufficient medical resources.

Naturally, it is important to note that this experience of temporal discontinuity generally associated with disasters and catastrophes, which produces fear, anguish, and unhappiness in people, is becoming increasingly frequent and widespread. It could be said that even political figures appear to be deliberately fostering this kind of fragmented perception of reality through public actions. As an example of the experience of discontinuity resulting from traumatic events such as disasters and catastrophes, consider the case of a person (let's assume she is a woman) who has lost her home, her husband, and all her possessions, and as a consequence needs to be relocated elsewhere. Certainly, this woman's life suffered because of the effects, for example, of Hurricane Katrina in New Orleans in 2005—a devastating rupture in her temporal connection between the present and the future. It could be said that depending on economic conditions, considering that disasters and catastrophes affect people from the upper and lower social classes differently, she could be thrown into much worse life circumstances that are very similar to previous stages of her life, and thus have to face challenges again that she had already overcome. This would significantly affect her self-esteem and could have negative psychological effects.

As another example of the experience of discontinuity as a result of traumatic events such as disasters and catastrophes, consider the case of a person (let's assume he is now a middle-aged man) who has lost his wife during the COVID-19 pandemic, which would certainly affect his quality of life and his sense of continuity in his life in relation to the future; he probably had plans for a life together with his wife that can no longer be realized. Depending on his financial circumstances, it may be the case that he may find it difficult to deal with all of life's challenges without the support of someone who has been with him for some years and shared these tasks with him. It may even be the case that she was the only person or one of the only people in his close family left, which would make it significantly difficult for him to rebuild his feeling of being at home in his own life again, and it would require a great effort to find a way in which his life would once again have a sense of continuity.

> We might ask, how much harm is enough? Why does time seem important in some instances (an earthquake) and not in others (a long-term drought)? And why are we much more likely to call harm in a confined spatial area a disaster (a plane crash, for example) and not describe as a disaster considerably more harm that takes place in a larger area over a longer period of time (the number of annual car accidents, for example)? (VOICE, 2015, p. 02)

I believe that Paul Voice's questions could be answered based on both the notion of temporal discontinuity and the prejudiced bias generally contained in the use of the term disaster in Western media, especially in North American media, when covering this type of event, as Naomi Zack (2009, p. 2) points out when dealing with the definition of disaster. Firstly, the prejudiced bias of media coverage tends to consider the death of a few people, if citizens of Western countries, a disaster and to give little or no coverage or even apply the term 'disaster' or 'catastrophe' to the death of a larger number of people from other parts of the world; secondly, the concept of temporal discontinuity can help to understand why a plane crash that kills two hundred people is a disaster or catastrophe and the death of thousands of people in traffic accidents over the course of a year is not. Likewise, earthquakes drastically affect the temporal continuity and often

even affect the spatial dimension (given that many people sometimes need to be displaced to other locations) of the lives and rational life plans of those directly affected by the event, whereas droughts, even when long-lasting, can be managed in the flow of normal life of a community through the displacement of water, by building large water reservoirs to store water from the rainy season for periods of scarcity, and thus the temporal continuity of the lives of those affected by drought can be minimized or even eliminated. Traffic accidents are part of people's normal lives and, except in more extreme cases, when they block roads for long periods, they generally do not affect the flow of normal life for people living in a given location. In other words, traffic accidents rarely produce temporary medium or long-term temporal discontinuity, at most the effect is very short-term. Plane accidents, in turn, are generally unexpected (given the low statistical rate of their occurrence compared to car accidents), affect the temporal continuity of the lives of several people, particularly those close to the victims (relatives, friends, etc.), and generally imply, in a single event, the modification of the normal lives of a large number of people in the work of rescuing the injured and the dead and in locating the wreckage of aircraft, among other things. Thus, it could be said that, even if damage is important in defining disasters and catastrophes, the magnitude of the damage needs to consider to what extent these damages affect the temporal continuity of normal life and the rational life plans of the affected people.

Paul Voice (2015) seeks to develop a normative concept of disaster based on Rawls' theory of justice and proposes a new definition of disaster that focuses on the negative consequences of events and the social institutions that establish and support the exercise of two capacities: the capacity for moral action and the capacity for effective citizenship (Voice, 2015, p. 3). Here it is important to remember that Rawls's justice as fairness does not apply directly to people, but to the basic structure of society, so a normative definition of disaster needs to focus on the negative effects of disasters and catastrophes on the main institutions of society (the basic structure of society) and on two human capacities: the capacity to have and develop a sense of justice (which is quite similar to the capacity for moral action proposed by Voice), given that disasters, if preparation has been inadequate, or if, even after adequate preparation, unforeseen circumstances have arisen in the preparation and that significantly burden the social and economic conditions of the affected society, the ability to follow normal morality may be impaired, especially if the scenario becomes one of extreme scarcity of resources basic to human survival. Unlike Voice's approach, the current study is not restricted to the conditions of citizenship in an active sense of political participation, but focuses on the ability to develop one's own rational life plans and the negative effects on these of temporal discontinuity resulting from disaster and catastrophe events. Voice also understands the capacity for moral action as related to the ability to undertake the rational plan of life in Rawlsian terms and that these events affect what he calls human dignity. Citizenship, in turn, would be dependent on the maintenance of political, economic, and legal institutions, often affected by disasters and catastrophes (VOICE, 2015, p.4).

Voice's approach is interesting because it establishes a normative criterion of diagnosis and reparation to determine when victims' claims are legitimate as matters of justice based on the effects of disasters on people's capabilities. In the words of Voice, "merely being harmed is not itself an injustice unless that harm affects one's dignity or citizenship capacities" (VOICE, 2015, p. 5). Furthermore, it seeks to outline what type of reparative actions or remedies would be appropriate for victims of disasters and catastrophes, namely, rebuilding institutions that serve

the dignity of citizens and that restore their moral capacities and their ability to exercise their rights (VOICE, 2015, p. 6).

Dumouchel in *Migration and Catastrophes* illustrates well how the same type of event can be considered as a disaster or a catastrophe depending on what human beings do in relation to it. He cites two nuclear accidents in Japan, one of which occurred in 2011 and is considered a catastrophe due to the site in Fukushima and as the people who lived in the areas affected by the accident were relocated from the place where they normally lived before the accident; therefore, as a type of lasting or permanent temporal discontinuity occurred in the rational plans of the lives of the affected people, the event is classified as a catastrophe. On the other hand, on September 30, 1991, a nuclear accident also occurred in Tokaimura, but since people were not relocated from the places they lived before the accident, the event is not classified as a catastrophe, but only as a disaster, because it meant a temporary discontinuity in the medium or short term. However, considered only from a technical point of view, the Tokaimura accident was as dangerous as the Fukushima accident and what seems to have produced the catastrophic character, the long-term or permanent discontinuity, was the human decision to displace the people who lived in the affected region in one case and not in the other case.

Adequate preparation as a strategy against temporal discontinuity

When discussing disaster and catastrophe scenarios, a shared view is that ordinary morality, the moral rules and virtues appropriate to normal everyday life, do not apply to these types of extreme scenarios. This is a topic addressed in Naomi Zack's now classic book, *Ethics for Disasters*, reflecting on ethics in disaster scenarios. The author offers a response—I would even venture to say—quite sophisticated, to this type of view. She argues, as a starting point for reflection, that ethics does not take a vacation in disaster situations, or at least, it is not appropriate to consider that the moral rules and moral virtues of our daily lives cease to be valid in this type of scenario. She continues that this kind of view can only make sense if it ignores the distinction between disaster preparedness and response, particularly when it assumes that preparation that is known to be inadequate is justified. Inadequate preparation, for example, would be preparation that assumes that the best thing to do in a disaster is to save as many people as possible, not everyone. The Titanic ship that sank in the North Atlantic Ocean set sail from port with fewer lifeboats than crew, so it was assumed in its preparations that it would not be possible to save all the people onboard if a disaster occurred. Being committed to the moral and justice rule that all people must be saved at the time of preparation requires setting sail with enough lifeboats for the entire crew, particularly considering that, as happened in this shipwreck, the poorest people were the last to, or could not access the lifeboats. This is therefore completely different from what a theory committed to a relevant social position (that of the least favored) proposes, as is the case with Rawlsian justice as fairness should prioritize the least favored and not the most favored.

It should be noted, however, that this does not mean that, even after adequate preparation, and for example when the number of lifeboats is sufficient to save all the people involved, that some people do not die as a consequence of a disaster, due to the occurrence of unforeseen circumstances in a natural disaster or as a result of human action. In other words, there may indeed be extreme disaster situations, especially catastrophes, in which the response will always be insufficient or inadequate, either to return to normal the lives of those affected (a scenario in

which it is possible to plan life because there is temporal continuity), or to save the lives of all those affected by the disaster or catastrophe event, or because the consequences of the event were so drastic and unforeseen that preparation proved insufficient to provide an adequate response. Because of this, the current study argues that paying attention to the distinction between disaster preparedness and response is a fruitful strategy against long-lasting, persistent, or permanent temporal discontinuity and may even mitigate the extent of short- and medium-term temporal discontinuity. If in a given location there is a high incidence of floods, hurricanes, earthquakes, for example, and the authorities do not take any kind of measures to minimize, mitigate, or even eliminate the damage, especially to the most vulnerable people, the least favored people in society, it is evident that, implicitly at least, there is no commitment on the part of the rulers to the moral rule that the lives of all human beings matter equally, and in some cases certain rulers have already expressed themselves publicly indifferent in relation to disasters such as pandemics. One thing is that, despite the best possible preparation at the time, a disaster or catastrophe could have totally unforeseen or beyond-control circumstances that make it necessary to do some kind of triage, and consequently fail to save all the people who can be saved. Another completely different thing is a government not considering it a priority to carry out preparation work to mitigate the effects of a disaster, for example, a flood, often for ideological political reasons (because it is against human rights, because the person is a science denier, because another opposition party advocates for better preparation and wants to remain in the opposition despite its responsibilities as a public administrator) and thereby deliberately causing the death of many of its citizens due to the need to carry out a triage and, in doing so, leaving the most vulnerable people, the least favored people in society, in the background. Any government that fails to pay due attention to disaster preparedness (for whatever reason) and therefore allows periods of temporal discontinuity to occur that significantly affect the ability of human beings under its responsibility to undertake their rational life projects, especially the least favored (the most vulnerable), disrespects the value of human life of those affected and may even be held criminally liable for the adverse effects resulting from its omissions. The "vulnerability analysis allows us to see human faults and/or neglect (...) when disasters are not results of a contingent, unmoral nature, but rather a defective, unjust culture" (LAUTA, 2020, p. 46). In these cases, disasters cannot be understood as simple cases of extreme necessity, which does not recognize law, to employ the idea of Kantian law of necessity, but can be examined by the judiciary and considered as acts of human negligence and omission. Another point worth emphasizing is that what Lauta is discussing is not the issue of increased crime during disaster events, but disasters themselves as crimes, and therefore susceptible to being attributed to responsible authorities or other people whose omission or negligence can be identified (LAUTA, 2020, p. 47). Furthermore, even though there is generally an increase in crime during disasters and catastrophes, in this case it is about the type of virtues that are expected of victims and people who carry out humanitarian actions during a disaster or catastrophe. When we are talking about the legal liability of political authorities, mainly before but also after disasters and catastrophes, we are dealing with crimes that can be committed by these authorities, crimes of liability for the bad consequences that people under the legal responsibility of these authorities are subjected to because of omissions and the negligence of the authorities.

Disasters and catastrophes as a matter of justice

In his conception of justice as fairness, the North American philosopher John Rawls starts from the assumption that human beings must choose a theory of justice in a hypothetical condition called the original position. He describes the parties or people responsible for choosing the principles as having two faculties or capacities: the capacity to develop a sense of justice and the capacity to develop a rational plan of life. The ability to have a sense of justice depends on the perception of the social institutions under which one lives as just, and the ability to develop a conception of life or a rational life plan depends on the provision of certain resources and the way in which the individual manages these resources over time.

Rawls argues that the question of identifying which principles of justice are most appropriate for ordering what he calls the basic structure of society (which includes the political constitution, the market system, and even the family, among other things) concerns a society that lives under conditions of moderate scarcity, those conditions that Hume called circumstances of justice. In other words, it is not a scenario of abundance of resources, since in this type of case there is no sense in having demands for justice, but it is also not a scenario of extreme scarcity, in which case there would be little to redistribute in response to the demands for justice. It is highly likely that the disaster and catastrophe scenarios discussed here will, in many cases, come close, or could even come close if adequate preparation for the disaster event has not been carried out, to scenarios of extreme scarcity of certain resources, since during a disaster event certain resources such as food, water, and other personal hygiene items may become less available or scarcer and may suffer significant price inflation due to the abrupt drop in supply and the maintenance of demand, if not to the expansion of demand, which may make their prices more prohibitive. Although disaster and catastrophe scenarios may be cases that deviate from the justice scenario, it is important to note that the present study understands, firstly, that, given the increase in human knowledge in predicting natural disasters or those produced by human acts, there is a proportional increase in the responsibility of authorities, especially political ones, in developing public policies in preparation, response, and even mitigation or elimination of the harmful effects of natural disasters or those produced by human acts on people's lives, especially the lives of the least favored people in society. It is necessary to consider here that the least favored people in society are those who suffer most from the harmful effects of natural disasters or disasters caused by human acts and are those who are most likely to suffer what we will call throughout the text long-term or permanent (temporal) discontinuities, and this phenomenon has negative consequences in relation to their ability to undertake their rational life plans. Second, since the ability to undertake a rational life plan is an assumption of a theory of justice for contemporary liberal democratic societies, and any conception of justice chosen to order the basic structure of society needs to take this into account, a just society must be considered one that contains a commitment to seeking to ensure that people do not suffer arbitrary temporal discontinuities that negatively affect their rational life plans. This planning, related to rational life plans, requires, among other things, temporal continuity, the ability to think of a logical sequence between past, present, and future in the lives of people in society. It requires the ability to consider with some clarity that the actions we performed in the past are part of a plan that is either under development in the present or in the process of being implemented in the near or even more distant future. Thirdly, contrary to what one might think from the assumption that justice only takes place in circumstances of moderate scarcity, in the same way that Zack (2009) defends in *Ethics for Disaster*, it is understood that ethics and justice

do not take a vacation in catastrophe and disaster scenarios and as a consequence of this, we do not start here from the assumption that ethical reflection on disasters and catastrophes needs to be identified with scenarios of lifeboat or runaway train ethics, in which the only relevant ethical decision is one based on a type of consequentialism regarding who should die and who should live. There is more philosophical work to be done in ethical reflection on disasters and catastrophes than simply applying some version of consequentialism to these types of cases to see which best fits our preexisting moral intuitions. It is important to note that some even question the role of moral intuitions in identifying moral theories, and one might also question the relevance of using such extreme cases as disaster ethics as a laboratory for identifying the best moral theory. Moral theories must first and foremost explain ethical conduct in normal times, not just ethical conduct in apocalyptic scenarios.

As the current study also addresses the question of whether a disaster puts ethics on vacation, since it affects people's lives in such a way that some theorists equate disaster scenarios with war scenarios and therefore argue that, as in a war, ethical values are not valid in disaster scenarios and therefore one could think of living in a situation in which justice might not make sense, therefore, the sense of justice would also not be promoted in a war scenario or in a disaster scenario. However, I believe that the greatest effects of disaster and catastrophe scenarios are not on the sense of justice, but rather on the rational life plans of people in society, just as I do not share the view that disasters and catastrophes are scenarios in which ethics goes on vacation and therefore I do not share the view that only one of the views of normative ethics is adequate for disasters, namely, some version of utilitarianism. Properly understood disaster ethics has room for conduct guided by deontological, consequentialist, and virtue-based ethics, although it is necessary to examine in greater detail which deontological rules apply to these types of cases, which version of consequentialism, and which versions of the virtues are appropriate both for people affected by disaster events and catastrophes and for people who provide assistance or help to those affected.

I believe that the issue of life plans is central to ethics in disasters and catastrophes because this type of event is marked, at least for a significant part of society, by producing some kind of temporal discontinuity in the lives of those affected, which can temporarily or permanently hinder the implementation of the rational life projects of people affected by these events. The notions of temporal continuity and discontinuity play an important role in establishing meaning in human life, since without belief in this continuity of time in relation to the future, all efforts in the present may lack meaning. This temporal discontinuity can occur temporarily or permanently. When a natural disaster (or the result of human action) of small or medium proportion occurs, there is usually only a short-term temporary discontinuity, but when a natural disaster (or the result of human action) of large proportion occurs, the temporal discontinuity can be long-lasting or irreversible, at least considering the temporal sequence (past, present, and future) in such a way that a person's life has meaning. In these cases, at best, the temporal sequence could be restarted after the occurrence of the event or even by going back to a moment in the temporal sequence much earlier than the moment in which the catastrophe or disaster occurred, when, for example, a person from a humble family background, after a long life trajectory and effort, manages to climb some steps of the social ladder and achieve a more dignified and valuable living condition, but suddenly finds himself back in a worse period of his life with less dignity because a disaster event destroyed his family, his home, his car, and the company he worked for.

There is a moral argument in favor of humanitarian aid to people affected by disasters and catastrophes, and a fiduciary responsibility has increasingly been consolidated through legal processes against political authorities and other people involved in the process of preparing for disasters and catastrophes when there is an omission or negligence. If we try here to somehow follow in the footsteps of Rawls's justice as fairness and apply it to ethics in disasters and catastrophes, it is important to note that he does not adhere to the principle of reparation as the sole criterion of justice (RAWLS, 1971, p. 101) and that the principle that Rawls applies to inequalities and comparisons between unequal positions is the difference principle that assigns an important role to a relevant social position, namely, the position of the least advantaged in society. Another aspect that Rawls employs in the discussion of the difference principle and that may be relevant to bring up here, is the assertion that natural facts are neither just nor unjust, but justice is found in the way social institutions deal with natural facts. As previously observed, the increase in human knowledge and the abandonment of metaphysical explanations for certain facts, such as disasters, which were once considered in philosophical reflection as a reason for a debate about divine goodness or evil among thinkers like Jean Jacques Rousseau and Voltaire (François Marie Arouet), are today considered as facts that can be explained either through natural causes or through causes related to human actions. Even when dealing with natural causes, today, with increased knowledge, public authorities can adopt measures to mitigate or eliminate the negative consequences of a significant portion of disaster events. Therefore, it becomes even more evident that the fact that a human being is placed in a situation of temporal discontinuity that affects the possibility of future development of their rational life plan influences the way key social institutions handle the process of disaster preparedness and response.

Neither only utilitarianism nor a case of moral turmoil

Another point I would like to highlight is that it is very common, when addressing ethics in disasters, to understand that it is about the application of renowned normative theories (such as Kantianism, utilitarianism, and virtue ethics) to scenarios similar to the runaway train, in which the only possibility is to choose between two or more equally bad scenarios, in which the consequences for human lives or even the lives of non-human animals need to go through a type of screening or selection regarding who should live or die, who should suffer this or that harm in the case of different scenarios causing physical or mental harm to different people in the ethical dilemma in question.

Something that is important to note is that cases of ethics in disasters were generally considered cases of exception to conventional morality and relatively rare, nevertheless, either because scientific knowledge has advanced and can predict with relative certainty the occurrence of certain types of disasters or natural catastrophes, or because human action has affected in such a way, through pollution of air, water, soil, etc., cases of disasters have become much more frequent in the media and in people's lives. This does not mean that our increased scientific knowledge has increased the incidence of these types of events, but at least it has made us more aware of their occurrence in the present or in the future. Because of this, it is also important to make efforts to promote the dissemination and broad understanding of scientific knowledge that serves as a backdrop for the early identification of disaster or catastrophe events, including fighting the spread of *fake news* and making it easier to explain and justify the public or private investments that are necessary to avoid

or at least mitigate the negative effects on people's lives that can result in temporary, persistent, or even permanent discontinuities in their life plans.

As already seen, disaster preparedness occurs before the event has happened, while disaster response occurs during, immediately after, or even when the event is imminent. The point is that insufficient preparation usually leads to an insufficient response, and with regard to the topic of this study, insufficient preparation and insufficient response can make the disruption of normal life caused by the event longer and more lasting than would be appropriate, or can even turn something that should be considered simply a disaster into a catastrophe.

I am now going to present two versions of the view that morality takes a vacation in catastrophe or disaster events: the first view understands that "once the threshold is crossed, morality ends." (SANDIN, 2020, p. 20), that is, once the scenario of discontinuity in people's lives is established, even if temporarily, morality loses its validity and the appropriate position is a type of political realism in which the strength of the strongest prevails and not the rules and moral virtues of everyday life. The second view understands that "when the threshold is crossed, deontology is simply replaced by consequentialism." (SANDIN, 2020, p. 20), that is, once the scenario of continuity in people's lives is established, even if temporarily, deontological moral rules and virtues lose their validity and the appropriate morality is some kind of consequentialism or utilitarianism. This kind of vision is what motivates the large number of analyses and studies of disaster ethics based on cases like the runaway train problem, in which the only alternative is to choose between bad scenarios; the only option left is to choose the lesser of two or more evils; human lives will be lost; it is simply a matter of performing a calculation of consequences to identify which of the alternatives results in the fewest number of people killed or injured. One consequence of this view seems to be that utilitarianism or some variant of consequentialism is the best ethical conception for extreme situations such as disasters and catastrophes, and that duty ethics and virtue ethics would have no role to play in disaster scenarios. However, it does not seem correct to claim that utilitarianism is the best conception of normative ethics for disasters and deontological rules and virtues still play a role in this type of scenario and the background realist view often serves as a poor justification for erroneous virtue models for catastrophe and disaster events.

Zack (2009, p. 50) asks: "Are disaster heroes and the virtues they exemplify different from the heroes and virtues of normal times?" After rejecting the heroic model of virtues based on the character Achilles from the Trojan War, she uses the examples of a father and his son in the film *The Road*, comparing them to the crew of the ship Endurance commanded by Captain Ernest Shackleton, and finally argues that the main virtues in cases of disaster are integrity and diligence, in contrast to the reckless bravery and ferocity characteristic of the demigod Achilles in the Trojan War. While acknowledging that these two virtues do not belong to the traditional catalog of virtue ethics like Aristotle's, she understands that they are the virtues necessary in conditions that people fear most, such as disasters and catastrophes. It can be said that it is very common in catastrophes and disasters, especially when assuming that catastrophes and disasters are similar to or should be treated in the same way as war, to argue that the appropriate virtue is that of warriors like Achilles in the Trojan War of Homer's Iliad—namely, the virtue of the fearless and invincible warrior in the face of the difficulties of war. In other words, virtue in a disaster or catastrophe would have to be a type of heroic virtue. However, this model of virtue cannot be appropriate because the vast majority of human beings could not achieve this level of skill in the

face of extreme events. This also does not mean that there should be no concern for virtue and moral rules in disasters and catastrophes.

Although Zack identifies two virtues as the most important in disasters and catastrophes, she doesn't differentiate who should possess these virtues. Instead, Löfquist does some work to determine which virtue applies to whom, by dividing people into two broad groups: those who are victims or affected by such events, and those who respond to or are responsible for providing humanitarian aid to victims or affected people. Löfquist understands that the quintessential virtue of people affected by or victims of disasters or catastrophes is resilience (LÖFQUIST, 2020, p. 207) and the virtue par excellence in relation to people who are responding to the suffering of those affected or victims of disasters or catastrophes is the ability to respond to the suffering of others, sometimes called beneficence, sometimes called benevolence, and sometimes called humanity, and it refers specifically to doing good to people who are in need in an event of disaster or catastrophe (LÖFQUIST, 2020, p. 208). Furthermore, contrary to the approximation of disaster scenarios to the condition of necessity or the right of necessity (ius necessitatis), which as Kant says in *The Metaphysics of Morals* "necessity has no law (*necessitas non habet legem*)" (MS, AA 08: 236), disasters and catastrophes are events in which moral principles and virtues still apply, since, as Kant says "there is no necessity that makes licit that which is inconsistent with law" (MS, AA 08: 236), although here we are dealing with ethics and political morality and not law in the strict sense.

Another factor that has caused a feeling of discontinuity between the past, present, and future in people's lives is related to new digital media and the crisis in democracy, due to the widespread dissemination of inaccurate or untrue news, popularly known as fake news. As Byung-Chul Han (2022, p. 36) states in *Infocracy - Digitalization and the Crisis of Democracy*, the way information is transmitted through new digital media causes "time to decay into a mere succession of specific presents. This is where information differs from narratives, which generate temporal continuity." Faced with this scenario in which we are bombarded with news that is sometimes true, sometimes inaccurate or false, and which creates a constant perception that we are on the verge of some case of temporal discontinuity, a feeling of hopelessness is produced among people and rational life plans are affected, since "actions need a horizon of meaning" (HAN, 2024, p. 13). However, I will not delve into the role of new digital media in producing a scenario of constant threat of temporal discontinuity in this study; I intend to dedicate another article to this topic in the future.

Rawls dealing with contingency: an extension to disaster ethics

Disaster and catastrophe scenarios would be classified in Rawlsian justice theory as some type of non-ideal theory, and in *The Law of Peoples* he differentiates between two types of non-ideal theory: 1. one that "deals with conditions of noncompliance, that is, with conditions in which certain regimes refuse to comply" to rules; and 2. one that "deals with unfavorable conditions, that is, with the conditions of societies whose historical, social, and economic circumstances make [it] (...) difficult, if not impossible" to achieve a just society (RAWLS, 1999, p. 5). Since disaster and catastrophe scenarios generally entail situations in which the supply of primary social goods and other consumer and subsistence goods may become scarcer or even unavailable for a certain period of time without external help from other parts of the country or even from other countries in the world, then at least momentarily, but in some cases for even longer periods, a society that

previously lived under conditions of moderate scarcity, which characterizes the circumstances of justice in justice as fairness, begins to live with unfavorable conditions and, therefore, would be a non-ideal theory scenario of the second type. Considering that the unfavorable conditions involved in disaster scenarios do not fail to affect the acquiescence to social norms of people in the society affected by the event, since there is normally a greater incidence of petty theft after disasters and catastrophes, it could be stated that disasters and catastrophes could also be classified, although they do not necessarily and always become this type of situation, as being in the first type of non-ideal theory mentioned above. I say "not necessarily and [not] always" because I understand that if public authorities fulfill their obligations in preparing for and responding to disaster events (at least, given that catastrophe cases imply a wider discontinuity), the conditions that make the sense of justice less effective in binding social norms during these types of events become less likely, or in many cases even non-existent.

I believe that some of Rawls's statements in *Justice as Fairness: A Restatement* (§ 51), as a strategy for responding to some of Amartya Sen's criticisms of justice as fairness, could shed light on what the Rawlsian position on disaster preparedness and response to disasters and catastrophes might be. The focus of the section is health care and how the circumstances in which people need health care can be considered as cases in which citizens temporarily fall below the minimum level of basic capabilities of a fully cooperative and normal member of society (RAWLS, 2001, pp. 171-2). Disasters and catastrophes, as we have observed, produce a type of temporary discontinuity in the short, medium, or long term, that affects the capabilities of citizens, since they affect or even take out of operation several important social institutions responsible for protecting the rights and duties of citizens, what Rawls called the basic structure of society. Furthermore, Rawls argues that justice as fairness can help only with the question of how to specify the just terms of cooperation between free and equal persons, but that it could be expanded to help with the question of justice between citizens who differ due to illness and accident, particularly by recognizing that citizens during their lifetime "may be seriously ill or suffer severe accidents from time to time" (RAWLS, 2001, p. 172).

It is necessary to pay attention to three aspects of the index of primary social goods that give a certain flexibility to the two principles of justice as fairness, adjusting to the differences between citizens of society related to health care and, we could add, related to the negative effects of temporal discontinuity in relation to rational life plans produced by disasters and catastrophes (RAWLS, 2001, p. 172). a. primary social goods "are not specified in full detail (...) [based on the available information] in the original position" (Rawls, 2001, p. 172); b. the primary social goods "of income and wealth should not be identified only with personal income and private wealth" (RAWLS, 2001, p. 172), but also include citizens as beneficiaries of a wide range of public goods and services whose provision ensures public health and which may include clean air and clean water, among other things; c. the primary goods index "is an index of l expectations of these goods over the course of a complete life" (RAWLS, 2001, p. 172).

Rawls states that from the conception of citizens as free and equal and the flexibility of primary social goods, two things are possible: a. "to estimate the urgency of different kinds of medical care" (...) and b. "to specify the relative priority of the claims of medical care and public health generally with respect to other social needs and demands" (RAWLS, 2001, p. 174). If we apply the same reasoning that Rawls applied to issues of medical care to issues related to ethics in

disasters, taking into account that temporal discontinuity affects citizens' capabilities (sense of justice and conception of good) in a similar way to what happens with illnesses and accidents that affect people's health, we could conclude that just as health treatments have "great urgency, more exactly, the urgency specified by the principle of fair equality of opportunity" (RAWLS, 2001, p. 174) actions by the State aimed at preparing for, responding to, and recovering and rebuilding the main social institutions of the basic structure of society damaged by disasters and catastrophes are also urgent and necessary.

Final considerations

Considering that Rawlsian justice as fairness, through its conception of equality that includes the principle of fair equality of opportunity and the principle of difference and is not committed to the principle of reparation, which assumes that all undeserved inequality is worthy of compensation, how can we justify the obligation of the State and society in the face of the condition of victims of disasters and catastrophes? The justification for the State's obligation towards people affected by disasters and catastrophes can be found, firstly, in a moral argument that disasters and catastrophes are events resulting from natural causes or human actions (most often human actions of people other than those affected by the event) that cause great commotion because they are undeserved by those affected and cause temporal discontinuity in the rational plans of the people affected; secondly, the negative effects of temporal discontinuity on the rational plans of life and even the sense of justice of the people affected because of the damage caused to the main institutions of the basic structure of society entail an obligation on the part of government authorities to prepare and create institutions specialized in preparing for and responding to disasters (natural and caused by human action). Furthermore, given the negative effects of the emergence of new media in spreading the feeling of fear due to the diversity of almost daily threats of the emergence of new temporal discontinuities, authorities also need to take serious measures in regulating new social media in order to combat and hold users and technology companies accountable for the dissemination of information that may contribute to the worsening of the negative scenario.

Considering that disasters and catastrophes generate circumstances of temporary, provisional, or permanent temporal discontinuity and that discontinuity significantly affects the ability of citizens of a society to undertake their rational life plans, what obligations should a State have in relation to this type of event? First, the State must take appropriate measures to avoid, mitigate, or minimize the adverse consequences that produce circumstances of temporal discontinuity in people's rational life plans, such as protecting, repairing, or rebuilding the social institutions of the basic structure of the affected society; second, the State needs to create systems for monitoring locations where disasters and catastrophes are likely to occur and create systems that warn people affected by the event that it is about to happen, as well as to train those affected on good disaster response practices, with the aim of significantly reducing the number of victims, if possible to zero; third, the State must direct available human and financial resources to assist disaster victims and, when necessary, temporarily allocate victims to schools, sports centers, hospitals, etc. until the victims are able to continue their rational life plans on their own by returning to their place of origin, when possible. However, when this is not possible, other measures are necessary to prevent victims from abruptly and drastically returning to previous stages of development of their rational life plans. Fourth, when possible and feasible, the State must rebuild the structures necessary for the normal development of life in the affected area in order to enable, as quickly as possible, the

return to the temporal continuity of the rational life plans of the affected people. Finally, the State needs to create instances of deliberation and preparation for disasters and catastrophes before the events occur in order to establish, through the public use of reason in the public sphere, how necessary it is to carry out some type of triage, in case the preparation proves insufficient and not all people can be saved, in order to avoid such criteria being established based on the emotions of the moment.

Bibliographic References

DUMOUCHEL, P. 2016. Catastrophes and Time: Catastrophes and Temporal Discontinuities. *Ritsumeikan Studies in Language and Culture*. Kyoto, v. 28, n.1, p. 193-202.

DUMOUCHEL, P. 2017. Migration and Catastrophes: Introduction. Kyoto, *Ritsumeikan Studies in Language and Culture,* v. 29, n. 2, p. 1-2.

DUMOUCHEL, P. 2021. Catastrophes and Inequalities. *Studi di Sociologia*, n. 4, p. 327-338.

HAN, B. 2022. *Infocracia:* Digitalização e a crise da democracia. Tradução de Gabriel S. Phillipson. Petrópolis: Editora Vozes.

HAN, B. 2024. *O espírito da esperança:* Contra a sociedade do medo. Tradução de Milton Camargo Mota. Petrópolis: Editora Vozes.

KANT, Immanuel. 2011. *A metafísica dos costumes*. Tradução de José Lamego. Lisboa: Fundação Calouste Gulbenkian.

SANDIN, P. 2020. Conceptualizations of Disaster in Philosophy. In: O'MATHÚNA, D. P.; GORDIJN, B. (Orgs.). *Disasters*: Core Concepts and Ethics Theories, Ohio: Springer.

LAUTA, K. C. 2020. Disasters and Responsibility: Normative Issues for Law Following Disasters. In: O'MATHÚNA, D. P; GORDIJN, B. (Orgs.) *Disasters*: Core Concepts and Ethics Theories, Ohio: Springer.

LOFQUIST, L. 2020. Virtue Ethics and Disasters. In: O'MATHÚNA, D. P; GORDIJN, V. D. B. (Orgs.) *Disasters*: Core Concepts and Ethics Theories, Ohio: Springer.

O'MATHÚNA, D.P.; GORDIJN, B. 2020. Conceptualizing and Assessing Disasters: An Introduction. In: O'MATHÚNA, D. P; GORDIJN, B. (Orgs.) *Disasters*: Core Concepts and Ethics Theories, Ohio: Springer.

RAWLS, J. 2001. *Justice as Fairness:* A Restatement. Cambridge: The Belknap Press of Harvard University Press.

RAWLS, J. 1999. *The Law of Peoples*: With "Idea of Public Reason Revisited". Cambridge: Harvard University Press, 1999.

VOICE, P. 2015. What Do Liberal Democratic States Owe the Victims of Disasters?: A Rawlsian Account. *Journal of Applied Philosophy*, v. 33, n. 4, p. 396-410.

ZACK, N. 2009. *Ethics for Disaster.* New York: Rowman & Littlefield.

O conceito de povos em Rawls frente aos desafios da irracionalidade social

The Rawlsian concept of peoples and the challenges of social irrationality

Julia Sichieri Moura[1]
Universidade Federal do Paraná (UFPR)
juliasmoura@gmail.com

Abstract: O conceito de povos presente no livro *O Direito dos Povos* de John Rawls está entre as ideias mais criticadas do terceiro livro que compõe a trilogia rawlsiana. Neste artigo buscarei apresentar uma justificativa desta ideia a partir da análise do conceito por Philip Pettit, demonstrando a coerência do conceito de povos no âmbito da obra em que ele aparece e assinalando o papel que o mesmo ocupa em um perspectiva mais geral na teoria de *justiça como equidade*. O artigo defende que a ontologia dos povos em Rawls, tal como interpretada por Pettit, fornece critérios para avaliar a ilegitimidade de grupos que agem de modo irracional, oferecendo um recurso conceitual relevante para debates contemporâneos sobre extremismo.

Keywords: cosmopolitismo; extremismo; Philip Pettit; povos; PREVENT; John Rawls.

Resumo: The concept of peoples presented in John Rawls' book *The Law of Peoples* is among the most criticized ideas in the third book of Rawls' trilogy. In this article, I will seek to justify this idea based on Philip Pettit's analysis of the concept, demonstrating the coherence of the concept of peoples within the scope of the work in which it appears and highlighting the role it plays in a more general perspective in the theory of *justice as fairness*. The article argues that Rawls' ontology of peoples, as interpreted by Pettit, provides criteria for assessing the illegitimacy of groups that act irrationally, offering a relevant conceptual resource for contemporary debates on extremism.

Palavras-chave: cosmopolitanism; extremism; Philip Pettit; peoples; PREVENT; John Rawls

[1] A autora agradece à Fundação Araucária pelo apoio à pesquisa realizada, viabilizada pelo projeto PROPAR.

Recebido em 1º de setembro de 2025. Aceito em 11 de dezembro de 2025.

I. Introdução

No artigo buscarei, através da análise do conceito rawlsiano de *povos* efetuada por Philip Pettit, relacionar a análise da ontologia dos povos (concepção de sociedade em Rawls) com os desafios contemporâneos de irracionalidade social, elaborando especialmente a ideia de grupos e a legitimidade dos mesmos em sociedades democráticas. Para tal, o artigo será desenvolvido em três momentos: (i) a exposição das críticas ao conceito de povos, (ii) a reconstrução da defesa de Pettit e (iii) uma proposta de leitura sobre irracionalidade social com base nesta discussão.

O texto *O Direito dos Povos* (1999) é possivelmente o texto que gerou maior polêmica da obra de John Rawls. Muitas ideias ali contidas foram rechaçadas: o anticosmopolismo rawlsiano em um contexto no qual dois dos principais autores de propostas distributivistas globais fundamentam suas ideias na teoria de Rawls, a crítica à proposta compreendida como minimalista de direitos humanos, a crítica de que há elementos imperialistas contidos na teoria e, também, da mesma não se posicionar de modo adequado temendo a crítica do imperialismo (MARTIN; REIDY, 2006, p. 10-11). Estes elementos não serão tratados neste artigo[2], que tem os seguintes objetivos: a) apresentar a defesa de Philip Pettit do conceito de povos na concepção de justiça proposta por Rawls e b) relacionar os elementos constitutivos da ideia de povo em Rawls com um critério avaliativo de grupos que se expressam e com irracionalidade social.

Houve dois momentos na literatura especializada após a publicação de *O Direito dos Povos*. O primeiro, com grande representação de textos e autores que eram diretamente ou indiretamente herdeiros de Rawls, pois se situavam no âmbito da mesma tradição (liberal e analítica) e efetuaram a primeira recepção da teoria internacional de Rawls. São autores que conduziram a leitura da justiça igualitária de Rawls para a esfera internacional através da teoria cosmopolita e que já estavam seguindo este caminho quando Rawls publicou seu texto. A maior parte dos textos desta literatura secundária é de tom extremamente crítico, com uma argumentação que visou assinalar problemas em *O Direito dos Povos* para apontar para o caminho mais apropriado para tratar tais questões deveria ter como foco o indivíduo e não como base a ideia, contestada, de povos. É possível que se demarque posteriormente um segundo momento de avaliação do texto, com vozes ainda críticas, mas contrabalanceadas por autores como Philip Pettit (2005; 2006), Catherine Audard (2006) e Samuel Freeman (2006), com interpretações em defesa de *O Direito dos Povos.*

Dentre estes, o objetivo aqui é desenvolver o argumento apresentado no artigo de Pettit (2006), o qual interpreta o conceito de povos no contexto da concepção de Rawls a respeito da natureza das sociedades, isto é, da ontologia social da teoria rawlsiana. Com este texto, Pettit visa demonstrar não só que o conceito de povos é compatível com a teoria de Rawls, mas também que a compreensão desta articulação consegue explicar o caráter anticosmopolita de sua teoria. Trata-se de um argumento com especial importância aqui, não só por estabelecer elementos que podem sustentar a ideia de continuidade e visão holística da teoria de *justiça como equidade*, mas principalmente por se tomar como ponto de partida da teoria de justiça internacional o conceito de povos, e não de indivíduos. Por este motivo, optou-se por reconstruir o argumento de Pettit de forma detalhada para uma possível conexão do mesmo com o desafio contemporâneo de lidar com formas de irracionalidade social (fanatismo, extremismo, intolerância).

[2] Já tratei dos mesmos na obra *Compreendendo a Utopia Realizável: Uma Defesa do Ideal de Justiça Distributiva da Teoria de John Rawls* (2019). Este artigo retoma e expande argumentos que apresentei no capítulo 2 do livro.

II. O conceito de povos: críticas e uma possível defesa

A crítica ao conceito de povos apresentada por Rawls em *O Direito dos Povos* foi tematizada diretamente através de leituras das filósofas Martha Nussbaum (2002) e Seyla Benhabib (2004). Ambas compartilham da compreensão de que o conceito não consegue abarcar as demandas de justiça distributiva – no caso de Nussbaum, o foco é no direito das mulheres; já a leitura de Benhabib visa demonstrar as insuficiências conceituais no que tange aos fluxos migratórios. Sem desconsiderar as diferenças nas críticas apresentadas, vale destacar também que tais críticas decorrem de uma concepção de direitos humanos mais abrangente que a apresentada por Rawls. O conceito de povos é, assim, conceito-chave da teoria internacional de Rawls porque dele decorre a ideia de tolerância e de minimalismo dos direitos humanos. Benhabib, não obstante reconhecer que Thomas Pogge e Charles Beitz deram passos mais largos do que Rawls no que tange à problemática de "justiça entre as fronteiras" (*justice across borders*), ainda considera que o princípio da diferença no plano internacional não é adequado para tratar dessa questão (BENHABIB, 2004, p. 1761).

A filósofa defende o direito à cidadania (right to membership) como um dos direitos humanos (BENHABIB, 2004, p. 1762). Já Nussbaum desenvolve esta ideia de forma mais contida no artigo em questão, ao afirmar que o caminho a ser tomado deve ser no sentido de se estabelecer tratados internacionais que reafirmem os direitos humanos já estabelecidos e trabalhar para que as outras nações do mundo os implementem (NUSSBAUM, 2002, p. 299)[3].

Benhabib reconhece que um dos principais objetivos de Rawls ao designar as partes de "povos" é evitar a leitura que o realismo estabeleceu no campo da teoria internacional, que define os Estados como os principais atores da esfera global. Estabelece, assim, o conceito de povos para designar e tentar definir os agentes mais apropriados moral e sociologicamente para as discussões de justiça no plano internacional (BENHABIB, 2004, p. 1764.). Trata-se de uma diferenciação importante para que *O Direito dos Povos* possa se estabelecer como uma teoria que não retome o modelo tradicional (realista) de soberania, principalmente no que tange às concepções de soberania interna, com relação às pessoas que estão inseridas nos Estados, e externa, de se declarar guerras (BENHABIB, 2004, p. 1764). Assim, com a definição de condições morais (isto é, o respeito aos princípios já elencados) para o reconhecimento da legitimidade soberana dos Estados-membros da *Sociedade dos Povos*[4], Rawls limita o alcance de argumentos na esfera internacional que sejam baseados somente na soberania.

Nesta mesma linha argumentativa, Kupfer define a teoria internacional de Rawls como um sistema de Estados-nação unitário com soberania limitada (KUPFER, 2000, p. 641). Tal posicionamento, vale ressaltar, é tido como positivo por Seyla Benhabib. O que se torna problemático, para a filósofa, é que não obstante a intenção de Rawls, a ideia de povos é imprecisa, o que faz com que a distinção objetivada pelo autor entre povos e Estado na prática seja difícil de verificar e acabe se tornando uma forma de nacionalismo (BENHABIB, 2004, p. 1765- 1767).

A análise de Benhabib decorre das características constitutivas do conceito de povos assinalado por Rawls, especialmente da constatação de que há uma incompatibilidade entre a crítica à

[3] A abordagem de Nussbaum é com base na teoria das "capacidades", que se aproxima do discurso dos direitos humanos e não deve ser lida como uma rival da mesma (NUSSBAUM, 2006, p.291)

[4] A "Sociedade dos Povos" é o termo usado para se referir a todos os povos que seguem os ideais e os princípios de *O Direito dos Povos* em suas relações mútuas (RAWLS, 2004, p. 3).

concepção de soberania proposta pelo autor e a proposta de se definir povos através de um regime constitutional-democrático sem que este tenha alguma forma de soberania territorial. Eis como Benhabib formula esta crítica, considerada para a filósofa como um dilema na teoria:

> This then creates a dilemma for Rawls's theory: Either he must assume that peoples who are united by "common sympathies," and "ruled by a just constitutional government," are territorially organized semi-sovereign units, which possess features very much like states, or he must give up the stipulation that peoples are already organized into certain forms of government. If he were to accept the latter option, Rawls may need to revert to viewing individuals rather than organized peoples as the privileged units of reasoning about international justice (BENHABIB, 2004, p. 1765).

Verifica-se, com o texto acima, que a autora questiona a própria possibilidade do realismo utópico de Rawls, pois o critério principal de Rawls continuaria sendo a unidade estatal, vinculada ao realismo. Cabe ainda ressaltar que Benhabib, com sua formação na teoria crítica e sociológica, criticará a formulação do conceito de povos, tal como ele é apresentado por Rawls, isto é, através da ideia de natureza moral (com uma compreensão holística dos valores e práticas que os definem). Afirmará que decorre de uma concepção há muito ultrapassada nas ciências sociais, pois não considera que povo também é constituído pelas inúmeras esferas sociais que o perpassam como gênero, classe, etnia e religião (BENHABIB, 2004, p. 1766).

Já Pettit afirmará que há três elementos que se destacam no anticosmopolitismo de Rawls, a saber: *o argumento doméstico*, que afirma que a justiça no âmbito doméstico de sociedades bem-ordenadas estabelece demandas substantivas para a sociedade e a responsabilidade da mesma por seus membros; *o argumento internacional negativo*, ou seja, a justiça não faz as mesmas demandas entre as sociedades bem-ordenadas; e por fim, *o argumento internacional positivo*, isto é, a asserção de que as demandas de justiça entre as sociedades bem-ordenadas surgem no contexto de auxílio para as sociedades que são vítimas de opressão.

O fio condutor do argumento de Pettit é que tais elementos, que fundamentam o posicionamento de Rawls, não decorrem do pragmatismo ou do entendimento de que o cosmopolitismo estabelece demandas excessivas, sendo portanto, utópico. Para o autor, o anticosmopolitismo de Rawls se fundamenta na ontologia do conceito povos (PETTIT, 2006, p. 40-41). Ao desenvolver esta ideia, Pettit consegue apontar para a relação e diferenciação entre os conceitos de indivíduo, grupo e povo no arcabouço conceitual rawlsiano.

Retomando as características afirmadas por Rawls como constitutivas dos povos (extensão, agência e pressupostos para que os mesmos sejam representados por seus governantes), Pettit avaliará de que modo as implicações das mesmas configuram a ontologia dos povos. Iniciando pelo conceito de extensão, Pettit retoma a classificação das sociedades estabelecida por Rawls, a qual vale retomar aqui:

> Proponho considerar cinco tipos de sociedades nacionais. A primeira são os *povos liberais razoáveis;* a segunda, *povos decentes* [...]. Em terceiro lugar, há *Estados* fora da lei e, em quarto, *sociedades sob condições desfavoráveis*. Finalmente, em quinto, temos as sociedades que são os *absolutismos benevolentes* (RAWLS, 2004, p. 4-5).

Pettit nota que Rawls não recorre à terminologia "povos" para referenciar os três tipos de sociedades que não são "bem-ordenadas", isto é, as três últimas. A relutância de Rawls em tratar os outros três tipos de sociedades como "povos" deve ser levada em conta e considerada na configuração da ontologia de povos. Ou seja, esta deve responder porque somente as sociedades bem-ordenadas são consideradas na abrangência do conceito de "povos" (PETTIT, 2006, p. 42).

Quanto à agência dos mesmos, Pettit afirma que na estrutura proposta por Rawls os povos

são caracterizados de forma semelhante à psicologia do agente individual, isto é, os mesmos possuem "motivos morais", possuem orgulho e um senso de honra (pela sua história, por exemplo), além de poderem respeitar e exigir respeito e reconhecimento. Com esta caracterização, os povos podem agir em três frentes: como seu próprio governo (na esfera constitucional, como autores da constituição, por exemplo), na esfera doméstica (com relação aos outros cidadãos) e na esfera internacional (com relação aos outros povos). Nos dois últimos casos, o povo agiria através do governo (PETTIT, 2006, p. 43). Tal vínculo entre povo e governo, Pettit esclarece ao tratar dos pressupostos para a representação do povo. Eis o que afirma o autor neste sentido:

> A people will exist as an agent on the domestic and international fronts, then, only if the government acts appropriately in its representative role, giving the people a voice and a presence on those fronts (PETTIT, 2006, p. 43).

Destaca-se, assim, o pressuposto de que para se representar o povo, é necessário que o governo atue de modo apropriado, isto é, que seja limitado pela concepção pública de justiça. Mais uma vez, Pettit demarca que esta ideia – de que tanto a sociedade liberal quanto a sociedade decente devem se fundamentar em uma concepção pública de justiça – tem uma consequência que não pode passar despercebida. Isto é, se um governo foi injusto na esfera doméstica, não será possível que se fale em representatividade do povo em suas ações. No entender de Pettit:

> This is a striking claim. Let the government be domestically unjust, Rawls suggests, and there will be no people present in its actions. The government will have to be seen as a body that acts only in its own name and, he would say, as a body that has no standing under the law of peoples. The norms that tell us how the government should behave in relation to its citizens are constitutive norms that determine what it is to represent the people, not regulative norms that merely instruct us on how representation is best pursued. Suppose a government breaches those norms through failing to behave with respect towards its citizens. In that case we might be tempted to say that while the government still represents it peoples, it represents them badly. *But Rawls speaks as if it does not represent a people at all. [...]. It usurps the position of the people* (PETTIT, 2006, p. 43 – itálico nosso).

Portanto, o entendimento de que o governo não representa o povo quando este age em desacordo com a concepção pública de justiça pode explicar o motivo pelo qual Rawls restringe o conceito de povo para sociedades bem-ordenadas. O esclarecimento do conceito de povo através dos argumentos apresentados não é, evidentemente, suficiente para estabelecer a ontologia dos povos em Rawls. Para tal, Pettit tem que dar um passo além, ele deve identificar (a) os componentes constitutivos dos povos, (b) caracterizar de que modo tais elementos se relacionam entre si, e também (c) com os outros grupamentos que podem ser denominados de povos na estrutura proposta.

Pettit assinalará, neste sentido, que o componente básico do povo é a pessoa natural. Logo, é este ponto de partida da construção da ontologia dos povos (PETTIT, 2006, p. 44 e ss.). É a partir da constatação de que a pessoa natural é o elemento básico do conceito de povo que Pettit rearticula a ideia de expansão, agência e pressupostos de representação para conseguir configurar a ontologia pretendida. O passo subsequente é caracterizar de que modo as pessoas naturais relacionam-se entre si. Para responder a esta questão, Pettit questiona a estrutura que é formada através das relações entre as pessoas naturais. No caso do conceito em questão, que, como se verificou, trata das sociedades bem-ordenadas, depreende-se que as relações entre as pessoas naturais deve resultar em sociedades que sejam bem-ordenadas por comungarem de "razões compartilhadas". Nessas sociedades, tais razões compartilhadas legitimam a estrutura de governo e se tornam o fundamento que legitima a força e as decisões coercitivas do mesmo (PETTIT, 2006, p. 44-45).

Torna-se possível, assim, que Pettit avance em seu argumento e afirme que as pessoas organizadas da forma descrita configuram mais do que uma união de pessoas; constituem um grupo provido de agência[5]. Esta etapa de seu argumento é central para se compreender a diferença entre indivíduo, grupo e povo. Isto porque as pessoas podem unir-se em grupos que não necessariamente tenham agência, pois para caracterizar um grupo-agente, três outras condições devem ser satisfeitas: a primeira é que o grupo tenha objetivos compartilhados; em segundo lugar, o grupo deve compartilhar de posicionamentos a respeito desses objetivos, como organizá-los e a possibilidade de revisar seu conteúdo, ou seja, deve se responsabilizar pelo ordenamento dos mesmos; em terceiro lugar, o filósofo estabelece que esses objetivos e a forma de alcançá-los devem ser formulados e buscados com alguma racionalidade (more or less rational), ou seja, na dinâmica do grupo-agente, é necessário que este tenha como apreender e rechaçar projetos irracionais (PETTIT, 2006, p. 46). Para Pettit, evidencia-se que estabelecer essas condições para a caracterização de um grupo-agente é determinar que estes consigam simular minimamente agentes individuais.

O grupo, então, consegue se caracterizar como povo na teoria de Rawls quando ele se organiza em função de uma estrutura básica bem-ordenada, a qual se mantém através da contínua interação entre um governo representativo (modo exógeno de organização do grupo) e cidadãos participativos (modo endógeno de organização do grupo). Para Pettit, estas condições são satisfeitas quando:

> The people-as-represented in government will meet the three conditions for group agency. It will act for the realization of certain ends; it will act under the guidance of a body of judgements that members authorize as common property; and it will display a modicum of rationality in how it holds and acts on those ends and judgments (PETTIT, 2006, p. 48).

Nesse contexto, assevera Pettit, o anticosmopolitismo de Rawls decorre do fato de que tal posicionamento é o único possível quando se considera a *natureza* dos povos bem-ordenados. Vale destacar as palavras do autor:

> The answer is, I think, that he sees his anti-cosmopolitan position as the only one that sits easily with the nature of well-ordered peoples. [...]My claim is that by his lights cosmopolitanism fails to reflect an understanding of the nature of peoples. It fails to reflect an understanding of just what sort of a thing a people is. (PETTIT, 2006, p. 49).

A ideia de que o posicionamento anticosmopolita de Rawls é a conclusão que mais se alinha à sua ontologia política já havia sido explorada por Pettit no artigo *Rawls's Political Ontology* (2005), texto no qual o autor apresenta a concepção de *"civicity"*, que também vale ser retomada aqui. No texto de 2005, Pettit tem como objetivo apontar para o fato de que perpassa a teoria de Rawls um pressuposto que abarca a forma como os indivíduos se relacionam mutuamente e as estruturas nas quais eles se encontram imersos – trata-se da ideia de *civicity*. Este termo definirá a posição intermediária ocupada por Rawls, que rejeita tanto o "singularismo político" (decorren-

[5] A ideia de *agente coletivo* tem sido objeto de pesquisa de Philip Pettit nos últimos anos, com o livro *Group Agency: The Possibility, Design, and Status of Corporate Agents* (2011) – publicado juntamente com Christian List – sistematizando e desenvolvendo tal conceito. Trata-se, assim, da defesa da ideia de que os grupos podem efetivamente ter agência, o que significa que pode ser possível, quando se considera determinado grupo como agente, determinar responsabilidade ao mesmo, assim como avaliar os modos como este se relaciona com outros grupos-agentes. Convém assinalar que o elemento de autonomia é central, pois para que um agente coletivo seja autônomo na leitura de Pettit e List, o mesmo deve desenvolver um tipo de racionalidade específica, como demonstram os autors em *"to display the rationality that agency requires, its attitudes cannot be a majoritarian or other equally simple function of the attitudes of its members. The group agent has to establish and evolve a mind that is not just a majoritarian or similar reflection of its members' minds; in effect, it has to develop a mind of its own. This gives rise to the kind of autonomy that we ascribe to group agents*." (PETTIT; LIST, 2011, p. 8).

te da teoria libertária) quanto a ideia de "solidarismo político" (que se origina do utilitarismo)[6]. Definindo *civicity* como a concepção de uma sociedade política cujos representantes e governo agem de acordo com os valores e pressupostos que emergem no debate público (PETTIT, 2005, p. 168), Pettit afirmará que a ideia de sociedade bem-ordenada é muito próxima desse conceito. Interessa notar que o argumento de Pettit considera a teoria de Rawls em sua totalidade e, nos dois textos que tratam do tema, há uma retomada dos elementos de *Uma Teoria de Justiça* e do *Liberalismo Político* para fundamentar sua leitura. Este é outro ponto que coloca Pettit como voz dissonante da maioria de intérpretes da teoria rawlsiana.

Em tais termos, a leitura de Pettit justifica o anticosmopolitismo de Rawls na própria lógica interna da teoria rawlsiana. Logo, retomando o *argumento doméstico*, Pettit assinala que as obrigações que decorrem da concepção de justiça no sistema de Rawls se originam não da ideia de humanidade, e sim da vida compartilhada que existe necessariamente em uma sociedade bem-ordenada, o que explicaria também o *argumento internacional negativo*, pois essas relações não se sustentam do mesmo modo no plano internacional (isto é, com apoio nas razões compartilhadas que configuram as sociedades bem-ordenadas). Já o *argumento internacional positivo* se sustenta na ideia de estrutura de povo como grupo-agente, isto é, que aja como indivíduos e assim possa se relacionar na configuração da segunda posição original com outros grupos.

III. A ontologia dos povos como critério avaliativo da irracionalidade social

Verifica-se assim, por um lado, tais críticas cosmopolitas (NUSSBAUM, 2002; BENHABIB, 2006) situam a ideia de povos como um elemento que pode perpetuar irracionalidades sociais (exclusão das mulheres, fechamento para as outras formações sociais, como etnia e grupo e movimentos como fluxos migratórios), por outro, há um caminho outro a se seguir a partir da defesa de Pettit do conceito, entendido pela ideia do grupo-agente, o qual demonstra a possibilidade de um critério avaliativo que revela a irracionalidade social e busca limitar o papel da mesma (ou mostrar sua ilegitimidade). Esta discussão vale ser retomada e desenvolvida com maior aprofundamento no âmbito dos debates atuais sobre fanatismo, extremismo e as insuficiências de uma teoria como a rawlsiana para tratar adequadamente destes fenômenos.

É precisamente aqui que a ontologia rawlsiana dos povos, como grupo-agente, se conecta com diagnósticos contemporâneos de irracionalidade social. Quanto ao fenômeno do fanatismo, Katsfanas (2019) afirma que a filosofia tradicional (Locke, Hume, Shaftesbury, Kant) em suas análises de fanatismo caracterizam este fenômeno como tripartite: 1) um comprometimento inabalável com um ideal, 2) uma rejeição de se sujeitar tal ideal (ou suas premissas) à crítica racional e 3) a pressuposição de uma sanção não racional para este ideal. A crítica de Katsfanas é a de que esta formulação de fanatismo possibilita que uma pessoa que pode ser pacífica e tolerante

[6] Eis como Pettit esclarece esta ideia: "Under the solidarist view, the individuals who constitute political society have relationships with one another of such a kind that they constitute a group agent, establishing a single system of belief and desire. Under the singularist alternative, as we may call it, there are no particular relationships, or none of any particular importance, that individuals in the same political society have to bear to one another. There may be no particular natural relationships between them, of course, such as those that bind members of the same family or tribe. While it is possible that individuals will have entered various contractual relationships with one another, or even with government authorities, it is not essential that they should have done this. For all that belonging to the same political society requires, people may relate to one another in just about any fashion; they may be as heterogeneous and disconnected as the set of individuals who live worldwide at the same latitude. The point is naturally expressed by saying that the political people, far from being a group agent of any kind, are a mere aggregate of separate subjects" (PETTIT, 2005, p. 162).

também seja uma fanática, seu argumento é o de que o que caracteriza uma pessoa fanática é 1) a adoção de determinados valores como sagrados, 2) a necessidade de tratar estes valores como incondicionais para preservar uma determinada forma de unidade psíquica, 3) o senso de que o status de tais valores estão ameaçados pela falta de uma aceitação geral dos mesmos e 4) a identificação com o grupo, o qual é definido por um comprometimento compartilhado se tais valores sagrados. Neste sentido, é possível que estes valores sejam criticados não por serem falsos mas por promoverem uma forma específica de patologia social.

Considerar a interpretação de povos tal como definida por Pettit é compatível com uma compreensão de que os grupos providos de agência na ontologia rawlsiana fornecem elementos de identificação e indicação de ilegitimidade na agência de movimentos fanáticos (os quais não poderiam ser considerados de grupos-agente pela definição elaborada por Rawls). Voltando à crítica de Benhabib, que estabelece que Rawls deve mudar para uma perspectiva "individual" ou "estatista" de sua concepção, verifica-se que os elementos que perpassam a teoria rawlsiana não colocam a questão nestes termos, pois trata-se de sempre se pensar a perspectiva adequada para se refletir sobre determinada questão. Neste sentido, em inúmeros trechos de sua obra, Rawls afirma a questão do posicionamento de perspectiva adequada frente à situação que se analisa. Destaco aqui o último parágrafo de *Uma Teoria de Justiça* quando trata da natureza hipotética da *Posição Original* e o motivo pelo qual devemos recorrer a este dispositivo hipotético. Rawls assinala que quando compreendendo a concepção desta ideia, podemos em qualquer momento olhar para o mundo social através da perspectiva adequada e, também, se trata de um ponto de vista objetivo pois expressa a nossa autonomia (RAWLS, 1971, p. 587). O resgate da ideia de povos neste texto visou demonstrar uma possível conexão entre um conceito pouco estudado até mesmo por intérpretes de Rawls e uma possibilidade de leitura do mesmo que o conecta com discussões contemporâneas sobre irracionalidade social.

Neste sentido ainda, retomando o argumento doméstico *civicity*, demonstra-se o papel importante de se conceber os valores da vida compartilhada que informam a concepção de justiça no sistema rawlsiano. Aprofundar e analisar esta interpretação considerando-se os desafios concretos colocados pelo extremismo e a necessidade de se pensar políticas de contra-radicalização pode ser caminho promissor para se compreender até que ponto a teoria de Rawls tem elementos adequados para tratar de questões políticas contemporâneas como estas. A literatura filosófica sobre extremismo e fanatismo são ainda incipientes se pensarmos as mesmas em relação com a literatura sobre tolerância, que tem lastro teórico e prático centenário. Não se trata aqui de desmerecer tais temáticas e sim considerar de que modo elementos conceituais que se estabeleceram na cultura pública e na literatura e prática sobre a tolerância podem complementar e se relacionam com estes desafios.

Cito, por fim, um exemplo que considero interessante para ilustrar esta possibilidade. Pensemos no debate sobre a estratégia de contraterrorisimo do Reino Unido denominada PREVENT[7] que visa prevenir a radicalização dos indivíduos para o terrorismo. Tal estratégia, criticada (entre outros motivos) por situar a questão do surgimento de extremismos como algo contagioso ou uma doença, subestima o quanto que as pessoas são radicalizadas por argumentos e narrativas que nutrem suas ideologias, justificativas e valores (CASSAM, 2022, p. 35). Destaca-se que a estratégia PREVENT também é criticada por ter aspectos racistas, islamofóbicos e por permitir

[7] URL da página oficial da PREVENT: https://www.counterterrorism.police.uk/what-we-do/prevent/.

a violação de direitos humanos.

Em artigo recente Bentley e Woodford (2023) afirmam que PREVENT assume premissas rawlsianas ao colocar os termos do debate e excluir *a priori* determinados segmentos da população do Reino Unido da conversa democrática. Melhor seria, segundo as autoras, considerar um modelo de contra-radicalização baseado na crítica de Stanley Cavell a Rawls, que estabelece as virtudes de responsividade de escuta e vontade de mudar pelas partes. Elementos como o esclarecimento do projeto rawlsiano e de conceitos como o de povos na sua elaboração podem não só mostrar possíveis equívocos nesta análise mas também revelar elementos teóricos da obra rawlsiana que fundamentam práticas mais inclusivas ao mesmo tempo que não possibilite a validação de grupos extremistas e fanáticos no debate. Uma visão socialmente vinculada da sociedade em Rawls oferece respostas que não são incompatíveis à crítica de Cavell, por exemplo. Desenvolver outros destes elementos em diálogo com estas demandas é caminho que não foi possível neste artigo, que teve como objetivo assinalar para tal possibilidade.

O objetivo que se buscou aqui foi demonstrar como a ontologia dos povos pode ser critério avaliativo que não recaia nos problemas de exclusão identificados por Bentley e Woodford (2023), por exemplo. Retomando a ideia das características do grupo-agente, temos que: 1) objetivos compartilhados; 2) possibilidade de revisão dos mesmos; 3) formulados com racionalidade e apreender e rechaçar propostas irracionais. Estas características possibilitam que se diferencie grupos-agentes de grupos irracionais. A discussão destes aspectos com a literatura do extremismo e do fanatismo demanda maior aprofundamento que não foi possível aqui. No entanto, a complementaridade temática foi possível de ser vislumbrada. Pensar a agência dos indivíduos e das comunidades (e as formas das relações que são cultivadas) tem sido foco de análises de interdisciplinar na área de PVE (Preventing Violent Extremism) com campos como psicologia, psiquiatria, educação, saúde pública, criminologia (STEPHENS; SIECKELINCK; BOUTELLIER, 2021, p. 347) mas ainda incipiente na área da filosofia política.

A articulação destes debates com a tese central do artigo, isto é, que ontologia dos povos é possível de ser compreendida como um critério avaliativo para distinguir grupos legítimos de grupos irracionais, ainda é incipiente pois demanda maiores esclarecimentos interpretativos sobre a teoria rawlsiana e a ideia de sociedade que decorre da mesma. No entanto, a crítica de que há uma base rawlsiana para estratégias como a da PREVENT fica nitidamente comprometida com um olhar mais atento ao texto rawlsiano e a concepção de sociedade no mesmo.

Neste sentido, o debate entre John Rawls e Stanley Cavell, citado pelas autoras em sua crítica à Rawls, é importante de ser investigado com maior aprofundamento. Seu fio condutor, que parte de fragmento textual no qual Rawls considera os planos de vida de um indivíduo racional (no âmbito da equidade e da justiça) como acima de qualquer de qualquer crítica ("beyond reporach"), assim como os infortúnios que decorrem do mesmo (RAWLS, 1971, p. 467), é interpretado por Cavell como uma fala dirigida aos outros no sentido de uma justificativa de sua própria (boa) sorte frente aos outros. Este mesmo termo contextualizado pode ser compreendido como uma relação do agente com ele mesmo para considerar um plano que não foi de acordo com o esperado (LEFEBVRE, 2023). No entanto, a interpretação de Cavell, que reconhece na metodologia da posição original e dos princípios de justiça articulados por Rawls um fundamento para se discutir a justiça em qualquer momento (ou seja, discutir quais caminhos devemos trilhar deve fazer parte de um diálogo contínuo na sociedade), afirma que a ideia de "above reproa-

ch" acaba comprometendo a proposta da justiça como equidade. Cavell irá propor uma forma de vida democrática que seria mais significativa ao manter uma abertura contínua baseada em três virtudes: escuta, responsividade para a diferença e vontade de mudar. Mais uma vez, cabe resgatar os elementos que constituem um grupo-agente: razões compartilhadas, capacidade de revisão e racionalidade (assim como capacidade de se identificar irracionalidades). Estas não são incompatíveis com os critérios de Cavell, ao contrário, há inúmeros elementos da teoria rawlsiana que podem representar estas ideias, talvez o principal deles seja o comprometimento da teoria com a reciprocidade. Lefebvre (2024) demonstra a proximidade entre as teses dos dois autores e afirma que é possível reconsiderar a relação entre ambos. Lefebvre sustenta que há um elemento da teoria de Rawls que é pouco tematizada por intérpretes de Rawls: isto é, que o cuidado da teoria com a formação subjetiva, argumentando que agir a partir da justiça como equidade demanda uma transformação pessoal (LEFEBVRE, 2024, p. 7). Este elemento é compatível com a virtude que representa a vontade de mudar. Há inúmeros outros pontos em que os dois autores seguem caminhos diferentes, mas assinalo aqui para os tópicos que são citados por Bentley e Woodford (2023) para diferenciar os dois modelos.

No artigo das autoras verifica-se a tentativa de se demonstrar dois modelos de diálogo, afirmando que PREVENT expressa um formato rawlsiano e o modelo de Cavell possibilitaria uma forma de democracia deliberativa mais adequada para as políticas de contra-radicalização. (BENTLEY; WOODFORD, 2023, p. 13). Neste sentido, como vantagens do modelo de Cavell, as autoras citam o contínuo reconhecimento da vida compartilhada e não de valores pré-estabelecidos (no caso, os valores britânicos), em segundo lugar, afirma-se que a democracia demanda igualdade para todas as pessoas e um comprometimento com a mesma (destaco aqui a ideia de reciprocidade em Rawls, que constitui dos elementos mais basilares da teoria identifica este comprometimento central) e, por fim, a ideia de confiança mútua. O objetivo de resgatar esta crítica é a de demonstrar que o modelo de contínuo diálogo e abertura proposto não é incompatível com a teoria de Rawls.

É discutível o ponto de partida interpretativo das autoras, o qual pode ser questionado por dois motivos correlatos: por um lado, recorrerem a um ponto de partida frágil, pois trata-se de uma crítica pontual à teoria de Rawls através da qual desconsideram a obra como um todo, consequentemente, seu principal argumento, isto é, que há um alinhamento do programa que criticam (PREVENT) com a teoria rawlsiana fica, também, prejudicado. Vale, porém, destacar o aspecto construtivo do artigo ao colocar no âmbito da teoria política e moral o importante papel avaliativo, de orientação e propositivo no âmbito de tais políticas. Neste sentido, este ponto merece maior elaboração por estudiosos de teorias morais e políticas. Uma atualização de diagnóstico das patologias sociais assim como das ferramentas conceituais que a filosofia política possui é um dos caminhos possíveis e promissores para esta agenda.

O artigo teve como objetivo analisar a ideia de *povos* a partir do livro *O Direito dos Povos* mas não restringir esta discussão aos argumentos sobre as consequências desta ideia para o posicionamento anticosmopolita de John Rawls. Ao apresentar a análise de Philip Pettit da *ontologia dos povos*, a ideia principal foi não só demonstrar de que modo esta ideia está coerente com o *corpus teórico rawlsiano* e mais do que isso, que a partir da mesma é possível que se pense a legitimidade de grupos (grupos-agente e a sua capacidade, entre outras, de formular planos de modo racional e tenha capacidade de rechaçar projetos irracionais) e a possibilidade da ideia de *povos* exercer

critério avaliativo frente às questões de irracionalidade social.

Rawls, cabe ressaltar, não recorre a terminologia de irracionalidade social. Muitos dos exemplos nos quais Rawls articula a ideia de irracional estão na justificativa de sua própria tarefa teórica e na tentativa de rechaçar interpretações incompatíveis com os ideais de *justiça como equidade*. Neste sentido Rawls afirma que não é irracional tentar uma teoria que seja melhor do que o intuicionismo e o utilitarismo (RAWLS, 1971, p. xvii); ao tratar do egoísmo (na caracterização das partes), Rawls afirma que o mesmo é consistente e não irracional, mas é incompatível com o ponto de vista moral (RAWLS, 1971, p. 117); sobre o véu da ignorância, poderia se questionar se não se trata de uma condição irracional (RAWLS, 1971, p. 120), respondendo que o acordo deve ser compreendido pelo ponto de vista de qualquer pessoa ao acaso, enfatizando o papel do ponto de vista adequado para a justiça. Nesta situação - de pessoas que se colocam como ponto de partida equitativo - discorre Rawls, doutrinas explicitamente racistas são irracionais, não são concepções morais e sim formas de supressão (RAWLS, 1971, p. 129). Quando trata do problema da inveja (§ 80, TJ), Rawls considera que a mesma pode *não ser irracional* e isto pode ocorrer quando alguém está em uma posição tão inferior pelo índice de bens primários que esta situação causa dano ao seu autorrespeito, podemos nos ressentir de sentir inveja em uma estrutura social que permite tal disparidade entre as pessoas.

De tal modo, quando se afirma a possibilidade do grupo-agente se orientar por planos racionais e de rechaçar planos irracionais, a ideia de ponto de vista adequado para aquela situação deve ser mobilizado. Destaco que Rawls finaliza *Uma Teoria de Justiça* com esta ideia. Afirma Rawls que a possibilidade de olharmos a situação humana não só pela perspectiva social mas também pela situação temporal nos coloca na "perspectiva da eternidade", pois não se trata de um ponto de vista transcendente e sim uma forma de pensamento e sentimento que pessoas racionais podem adotar no mundo. Quando fazemos isso, qualquer que seja a geração, é possível conectar as diferentes perspectivas individuais em um esquema, chegando, assim, a princípios regulativos que podem ser afirmados por todos pelo seu próprio ponto de vista (RAWLS, 1971, p. 514). Em outras palavras, o que constitui a irracionalidade muitas vezes não é dado de antemão, o recurso metodológico da posição original e do véu da ignorância são (entre outros conceitos da teoria de Rawls) formas de se situar de modo avaliativo e normativo: na esfera doméstica, internacional e mesmo intergeracional. São aspectos da teoria que demandam maior análise e que podem fundamentar projetos teóricos de teoria não-ideal direcionados a pensar sobre os desafios da irracionalidade social em suas diferentes dimensões.

Referências

AUDARD, C. 2006. Cultural Imperialism and Democratic Peace. In: MARTIN, R. (Ed.); REIDY, D. A. (Ed.). *Rawls's Law of Peoples*. 3. ed. Malden: Blackwell.

BENHABIB, S. 2004. The Law of Peoples, Distributive Justice and Migrations. *Fordham Law Review*, v. 72, p. 1761-1787.

BENTLEY, M.; WOODFORD, C. 2023. Above Reproach: Rawls, Cavell, and Emersonian Conversation as a New Model for Democratic Counter-Radicalisation Policy. *International Political Sociology*, [*s.l.*], v. 17, p. 1-18.

CASSAM, Q. 2022. *Extremism*: a philosophical analysis. New York; Routledge.

FREEMAN, S. 2006. Distributive Justice and The Law of Peoples. In: MARTIN, R. (Ed.); REIDY, D. A. (Ed.). *Rawls's Law of Peoples*. 3. ed. Malden: Blackwell.

KATSAFANAS, P. 2019. Fanaticism and Sacred Values. *Philosophers' Imprint*, v. 19. n. 17, p. 1-20.

KUPFER, A. 2000. Rawlsian Global Justice: Beyond the Law of Peoples to a Cosmopolitan Law of Persons. *Political Theory*, [*s.l*], v. 28, n. 5, p. 640-674.

LEFEBVRE, A. 2024. Stanley Cavell, John Rawls and moral perfectionism in liberal democracy. *European Journal of Political Theory,* [*s.l*], v. 25 , n. 1, p. 50-69.

MARTIN, R.; REIDY, D. A. 2006. Introduction: Reading Rawls's Law of Peoples. In: MARTIN, R. (Ed.); REIDY, D. A. (Ed.). *Rawls's Law of Peoples*. 3. ed. Malden: Blackwell.

MOURA, Julia. 2019. *Compreendendo a Utopia Realizável*: uma defesa do ideal de justiça distributiva da teoria de John Rawls. Rio de Janeiro: Lumen Juris.

NUSSBAUM, M. 2002. Women and The Law of Peoples. *Politics, Philosophy & Economy,* [*s.l*], v. 1, p. 283-306.

PETTIT, P; LIST, C. 2011. *Group Agency*: The Possibility, Design and Status of Corporate Agents. Oxford: Oxford University Press.

PETTIT, P. 2010. A Republican Law of Peoples. *European Journal of Political Theory*, [*s.l*], v. 9, p. 70-94.

PETTIT, P. 2005. Rawls's Political Ontology. *Politics, Philosophy & Economy,* [*s.l*], v. 6, n. 4(2), p. 1470-1594.

PHILIP, P. 2006. Rawls's Peoples. In: MARTIN, R. (Ed.); REIDY, D. A. (Ed.). *Rawls's Law of Peoples*. 3. ed. Malden: Blackwell.

RAWLS, J. 2004. *O Direito dos Povos.* 3. ed. São Paulo: Martins Fontes.

RAWLS, J. 1971. *Theory of Justice*. Cambridge: Belknap Press of Harvard University.

STEPHENS, W.; SIECKELINCK, S.; BOUTELLIER, H. 2021. Preventing Violent Extremism: a Review of the Literature. *Studies in Conflict & Terrorism*, [*s.l*], v. 44, n. 4, p. 346-361.

“Eliminar o pior é mais humano do que buscar o bem”: problemas de uma “práxis otimista” e esboço de emancipação “anti-otimista” a partir da Teoria Crítica tardia de Horkheimer

“Eliminating the worst is more human than seeking the good”: problems of an “optimistic praxis” and layout of an anti-optimistic emancipation based on Horkheimer’s late Critical Theory

Vilmar Debona[1]
Universidade Federal de Santa Catarina (UFSC)
debonavilmar@gmail.com

Abstract: O artigo parte da acusação de Habermas de que Horkheimer teria sido improdutivo em sua fase tardia, fazendo as tarefas da Teoria Crítica dependerem da teologia e se tornando “um filósofo da história demasiado negativista, um crítico da razão demasiado radical”. Diante disso, procura delinear, em especial a partir de uma análise de *Notizen* (*Apontamentos*) como produção derradeira de Horkheimer que Habermas despreza por não ser sistemática, conteúdos do que proponho chamar de *elementos de emancipação anti-otimista*. O objetivo principal do estudo é, assim, mostrar em que consistem esses elementos, que não indicariam apenas práticas de solidariedade em sentido amplo e como conceito negativo – estendido, inclusive, à esfera de uma “política negativa” –, mas performariam tudo o que pode caber em uma práxis que, conforme lemos em um dos fragmentos de *Notizen*, não pretende “buscar o bem”, mas “eliminar o pior”. Para tanto, sugiro que é preciso considerar a controversa presença do par pessimismo-otimismo em Horkheimer, que nem sempre é empregado em sentido estritamente schopenhaueriano; e, em particular, como se pode compreender a recomendação do fundador da Escola de Frankfurt de um pessimismo teórico complementado por um otimismo prático ou por uma “práxis otimista”.

Keywords: Max Horkheimer; negação; pessimismo; práxis emancipatória; Teoria Crítica.

Resumo: The article starts from Habermas’s accusation that Horkheimer had been unproductive in his late phase, making the tasks of Critical Theory depend on theology and becoming “too negativistic a philosopher of history, too radical a critic of reason”. With this, it seeks to outline, especially based on an analysis of *Notizen (Notes)* as Horkheimer’s final production that Habermas despises for not being systematic, contents of what I propose to call *elements of anti-optimistic emancipation*. The main objective of the study is, therefore, to show what these elements consist of, which would not only indicate practices of solidarity in a broad sense and as a negative concept – extended, even, to the sphere of a “negative policy” –, but would perform everything that can fit into a praxis that, as we read in one of the fragments of *Notizen*, does not intend to “seek the good”, but “eliminate the worst”. To this end, I propose that it is necessary to consider the controversial presence of the pair pessimism-optimism in Horkheimer, which is not always used in a strictly Schopenhauerian sense; and, in particular, how one can understand the recommendation of the founder of the Frankfurt School of a theoretical pessimism complemented by a practical optimism or by an “optimistic praxis”.

Palavras-chave: Max Horkheimer; negation; pessimism; praxis; Critical Theory.

[1] Este trabalho recebeu apoio da CAPES por meio de bolsa CAPES/Print da UFSC (processo 88887.912532/2023-00) e do CNPq por meio de bolsa de Produtividade em Pesquisa/PQ-C (processo 311785/2025-5). O artigo expõe resultados parciais de uma pesquisa de pós-doutorado desenvolvida entre 2024 e 2025, em parte na Goethe-Universität Frankfurt, sob supervisão do Professor Christoph Menke, e em parte na Universidade de São Paulo, sob supervisão da Professora Olgária Matos.

Recebido em 1º de maio de 2025. Aceito em 11 de dezembro de 2025.

Para Delamar Volpato Dutra

I

Jürgen Habermas, em um capítulo de livro de 1986 intitulado *Observações sobre o desenvolvimento da obra de Max Horkheimer*[2], acusou com termos fortes uma suposta falta de Horkheimer por não ter elaborado qualquer projeto novo de Teoria Crítica após seu retorno do exílio estadunidense. Acusa-o, também, de ter cedido a uma série de "desesperanças" desde que deixou a Alemanha devido ao avanço do nazismo. Não chega a usar o termo "pessimismo" para reprovar essa suposta "resignação"[3] de seu ex-diretor no Instituto de Pesquisa Social, mas classifica-o como um adepto de Nietzsche e como alguém que se tornou "um filósofo da história *demasiado negativista*, um crítico da razão demasiado radical" (HABERMAS, 2007, p. 291, grifo meu). Para Habermas, a produtividade científica de Horkheimer a ser levada a sério, a "substância da sua obra", estaria circunscrita às contribuições anteriores ao fim da Guerra, sendo a dos anos 1930 os artigos publicados na Revista do Instituto. Nos trabalhos em colaboração com Adorno entre 1941 e 1944, Habermas enxerga que "terminou por se completar a viragem para uma filosofia *negativista* da história" (idem, p. 275, grifo meu). Já a produção das décadas seguintes, em especial as *Notizen*, apontamentos do período entre 1949 e 1969, seriam meros registros diários desconexos, "crivados de contradições", um "disparate" (idem, p. 276) que, diferente de Adorno, escancara a impossibilidade de qualquer dialética, nem mesmo de uma dialética negativa.

Como (não) fez também alguns anos antes, em sua obra magna *Teoria da ação comunicativa*, Habermas não reconhece que o progressivo afastamento de Horkheimer em relação a ideais originários da Teoria Crítica, sobretudo de teor marxista não-ortodoxo nos termos do programa de um materialismo interdisciplinar, se dá pari passu a uma cada vez mais significativa leitura crítica de Schopenhauer e do pessimismo schopenhaueriano, marcantes já na sua adolescência; e não apenas de Nietzsche, nem apenas do aceite do diagnóstico de Weber com uma radicalização da reificação lukácsiana (CHIARELLO, 2001, p. 193). Sabemos por depoimentos do próprio Horkheimer que foi Pollock quem lhe apresentou Schopenhauer durante viagens conjuntas para estudos quando muito jovens. Não deixa de ser significativo, aliás, que na biografia intelectual de ambos – e, portanto, na história do próprio Instituto de Pesquisa Social e dos inícios da Teoria Crítica frankfurtiana – Pollock não tenha sido apenas o economista que reiteradamente fornecia as bases de economia política para a renovação de diagnósticos das contradições do capitalismo aos teóricos do Círculo, em Frankfurt ou nos Estados Unidos, do "capitalismo monopolista" ao "capitalismo de Estado" ou "capitalismo administrado". Foi o mesmo Pollock quem, num quarto de

[2] No original alemão, *Bemerkungen zur Entwicklungsgeschichte des Horkheimerschen Werkes*, publicado na coletânea organizada por A. Schmidt & N. Altwicker, *Max Horkheimer Heute: Werk und Wirkung*, Frankfurt a.M, Fischer Verlag, 1986, pp. 163-179, traduzido para o português por Maurício Chiarello na Revista *Educação e Filosofia*, Uberlândia, v. 21, n. 42, p. 273-293, jul./dez. 2007.

[3] "O fato de Horkheimer recorrer à teologia efetivamente, e não apenas hipoteticamente, resulta da ameaça de ruína de seus próprios fundamentos, uma vez que não somente a filosofia da história perdeu sua base histórica, mas também se radicalizou a crítica total da razão – o Horkheimer tardio não quer, é certo, *resignar-se* a isto, mas não vê nenhuma outra saída" (HABERMAS, 2007, p. 290 grifo meu).

estudantes, jogou na cama de Horkheimer os *Aforismos para a sabedoria de vida* de Schopenhauer, dizendo: "Você pode se interessar por isso"[4].

Como também sabemos, Habermas empreendeu o que, segundo ele, Horkheimer teria sido incapaz: elaborou um novo projeto emancipatório em torno da ideia de "ação comunicativa", no qual a razão instrumental, frente a cujo poderio avassalador e instransponível Horkheimer supostamente ficara inerte, deixaria de ter a última palavra em desfavor da razão comunicativa. O déficit normativo do qual o ex-assistente de Adorno acusara amplamente a Escola de Frankfurt[5] teria sido, dessa forma, superado, e os mais diversos ramos e novos representantes da Teoria Crítica surgiriam no bojo de fundamentações das amplas lutas contemporâneas por reconhecimento e direitos no âmbito das democracias.

Considerando esse contexto, o presente artigo não pretende adentrar nem na Teoria Crítica habermasiana propriamente dita nem na imensidão dos desdobramentos dela nas atualidades da Teoria Crítica. Apenas assume o referido texto de Habermas como fonte altamente didática para localizar o teor de acusações feitas de forma ampla e reiterada pela posteridade em detrimento de seu fundador Horkheimer, assim como em vista de ponderar, a partir de textos pouco debatidos da última fase de produção do autor de *Crítica da razão instrumental*, em que medida as referidas acusações não procedem inteiramente. Afinal, será mesmo que o Horkheimer tardio não teria elaborado absolutamente nenhuma tese ou hipótese, nem mesmo em forma de esboço, para o presente e para o futuro da Teoria Crítica? As páginas que seguem pretendem elucidar, então, quais seriam algumas das direções que Horkheimer teria indicado nesse período acusado de improdutivo para ulteriores desenvolvimentos da crítica. Em especial a partir de uma análise de fragmentos das *Notizen*, produção derradeira de Horkheimer que Habermas despreza como insignificante por não ser sistemática, será possível notar a ideia do que proponho chamar de *elementos de emancipação anti-otimista*. Esses elementos não indicariam apenas práticas de solidariedade em sentido amplo e como conceito negativo – estendido, inclusive, à esfera de uma "*política* negativa" –, mas performariam tudo o que pode caber em uma práxis que não pretende "buscar o bem", mas "eliminar o pior" (HORKHEIMER, 2008, p. 261). Há, nesse contexto, conteúdos que extrapolam a defesa de uma *teologia* negativa no âmbito do "anseio pelo inteiramente Outro" (HORKHEIMER, 2024b), esse mote amplamente confundido por parte da recepção dos escritos tardios de Horkheimer – talvez por influência direta de Habermas? – como queda em mera religiosidade positiva e dogmática[6].

[4] Em uma de suas últimas entrevistas, em 1972, Horkheimer conversou com Gerhard Rein: "Eu estava com Pollock, e Pollock veio ao meu quarto numa noite e jogou um pequeno livreto na minha cama - acho que era um livreto da Reclam - com um texto de Schopenhauer, e disse: 'Você pode se interessar por isso'. E ele tinha razão, eu fiquei tão interessado que li Schopenhauer intensamente. Então fui para Schopenhauer muito antes de terminar o 'Ensino Médio'" (HORKHEIMER, 2024b, p. 452, trad. minha). Em abril de 1968, no Prefácio à reedição (com o título *Teoria Crítica: uma documentação*) de um conjunto de textos da década de 1930 publicados na *Zeitschrift für Sozialforschung*, Horkheimer já tinha registrado o seguinte: "O pessimismo metafísico, momento implícito em todo pensamento genuinamente materialista, me foi familiar desde sempre. À obra de Schopenhauer devo meu primeiro contato com a filosofia; a relação com a doutrina de Hegel e de Marx, o desejo de compreender e de mudar a realidade social não comprometeram, apesar do contraste político, minha experiência com a sua filosofia" (HORKHEIMER, 2015a, p. 4).

[5] "Com a presente investigação [de *Teoria da ação comunicativa*], pretendo introduzir uma teoria da ação comunicativa que esclareça os fundamentos normativos de uma teoria crítica da sociedade. A teoria da ação comunicativa deve oferecer uma alternativa à filosofia da história a que esteve atada ainda a teoria crítica mais antiga e que se tornou insustentável" (HABERMAS, 2022, p. 592-593).

[6] É possível afirmar que parte significativa da questão se deva à falsa ideia de que Horkheimer teria simplesmente arruinado a crítica social por ter se tornado um schopenhaueriano resignado, um "Schopenhauer cristão", comprometendo o tipo de práxis que outrora fora uma das mais potentes atualizações de um materialismo marxista não ortodoxo. As "saídas desencantadas" que

Contudo, os elementos emancipatórios que se encontram latentes em artigos, fragmentos e entrevistas da última Teoria Crítica horkheimeriana podem se apresentar enevoados na letra do autor devido a especificidades de empregos nem sempre por inteiro consequentes do par pessimismo-otimismo, dado que "pessimismo" é comumente associado a "desesperança" pelo *common sens*e ou pela própria Filosofia, nesse caso com a particularidade dos novos fracassos e impotências da razão daquele período histórico. Pois se é em sentido schopenhaueriano que os termos são empregados – sendo o schopenhauerianismo de Horkheimer antimetafísico, antiquietista (FAZIO, 2023b) e antifinalista (MATOS, 1989), que por isso mesmo teria de ser imune a qualquer otimismo –, como então poderia proceder a reivindicação feita pelo pensador, nesses mesmos textos da fase tardia, de um "otimismo prático" para a ação emancipatória? É seguro afirmar que a semântica original de "pessimismo", lida de um modo específico por Horkheimer, é aquela encontrada na obra de Schopenhauer, considerado o primeiro sistematizador do *pessimismo filosófico moderno* (PLÜMACHER, 1884), crítica metafísica aos otimismos clássicos do "melhor dos mundos possíveis" que foi interpretada e dinamizada "hereticamente" pelo frankfurtiano como crítica social. Com efeito, já é mais do que consolidado o pressuposto de que vieram dessa tradição, que ombreia as tradições idealista e marxista como fontes da Escola de Frankfurt, as inspirações horkheimerianas para um pessimismo crítico e não-conformista[7]; ou, de outro modo, conforme elaborou Alfred Schmidt (1977), como fonte para o mal metafísico (Schopenhauer) na condição de complemento à fonte para o mal físico (Marx). Um tal pessimismo, no entanto, não teria de ser assegurado até o fim e em todos as esferas (na teórica e na prática), em vez de se transformar em uma "práxis otimista" comprometedora das premissas anteriores?

Para investigar essas questões, proponho deslocar o foco do Horkheimer tardio da conhecida e hipervalorizada entrevista sobre o "anseio pelo inteiramente Outro" para alguns escritos do mesmo período, em especial para as referidas *Notizen*, material extremamente rico em teses e críticas. Foi uma recomendação de Alfred Schmidt a de ser possível rebater a partir desse material as avaliações distorcidas de que, pelo fato de Horkheimer ter dedicado não pouco espaço à religião e à teologia em aspectos de seu pensamento tardio, a conclusão apressada seria a de que o fundador da Teoria Crítica teria traído a obra de toda a sua vida na religiosidade positiva: "[...] as *Notizen*, mais do que certas entrevistas do último Horkheimer", podem ser indicadas como material que impede "os erros mais grosseiros de avaliação" (SCHMIDT, 1977, p. 115, trad. minha). Esse procedimento metodológico permitirá notar que o problema central se refere a pelo menos três facetas de uma mesma questão: ao conceito de "negativo" como definidor das tarefas

Horkheimer vislumbrou e procurou elaborar em relação aos horrores do nazifascismo e ao que se seguiu a tais horrores incluíam a busca por um (mundo) "inteiramente Outro" como forma de não aceitar que a última palavra fosse do horror, da opressão ou do "mundo administrado" – em suma, do existente. Essa busca, por poder incluir *momentos* de uma teologia *negativa*, permitiu, embora não de forma direta por Habermas, a acusação de uma suposta defesa de Horkheimer de uma mera religiosidade dogmática; e como supunha também a descrença em qualquer projeto emancipatório afirmativo e robusto de sociedade, atraiu a acusação fácil de um "Horkheimer pessimista", sem preocupação com a definição de qual *pessimismo* se trataria. Com impacto pejorativo direto para a fortuna crítica desse pensador na posteridade, esse reducionismo concentrado em uma tese exposta por ele basicamente em uma entrevista (a do "anseio pelo inteiramente Outro"), é altamente prejudicial, pois impede ou dificulta a apreensão de muitas contribuições significativas que o pensador pode dispor, inclusive para os mais variados ramos e propostas que a própria Teoria Crítica elabora atualmente.

[7] O presente artigo pressupõe a inegável influência direta e constante de Schopenhauer em Horkheimer, sem se ocupar com os termos específicos, os períodos e os diversos estatutos filosóficos dessa influência. Considero esse assunto, em parte porque reiterado pelo próprio fundador da Teoria Crítica, como relativamente bem consolidado pela literatura especializada. Para notar esse consenso, cf. SCHMIDT (1977), VEAUTHIER (1988) e FAZIO (2023b).

da Teoria Crítica, à relação do conceito de "negativo" com o emprego do termo "pessimismo", e à problemática (re)definição de uma práxis como um "otimismo prático".

II

No seu referido texto de 1986, Habermas poderia até ser compreendido como voz dissonante ao ainda se impressionar com as desesperanças de Horkheimer a partir da segunda metade da década de 1940 em relação a grandes feitos históricos para a emancipação humana. O conhecido problema da virada da Teoria Crítica dos anos 1940 em relação aos anos 1930 culminaria na aporia da *Dialética do Esclarecimento*, qual seja, a autodestruição do esclarecimento[8], e na busca por saber "por que a humanidade, em vez de entrar em um estado verdadeiramente humano, está se afundando em uma nova espécie de barbárie" (ADORNO; HORKHEIMER, 1985, p. 9). Habermas enxerga o enfretamento da questão, notadamente por parte de Horkheimer e nem tanto por parte de Adorno, em termos de retirada das tarefas próprias da "teoria histórico-social" e de repasse das mesmas para uma crítica radical da razão que se apega à denúncia de um parentesco íntimo entre razão e dominação. Isso teria consistido em esvair a "esperança numa tensão dialética intrínseca ao processo histórico" (HABERMAS, 2007, p. 280), bem como em abandonar as esperanças de que a razão instrumental, por ser apenas um produto derivado da época burguesa, seria superável numa formação pós-burguesa da sociedade, mediante uma "razão substancial normativa". Em uma palavra, Adorno e Horkheimer teriam submergido em um tipo de pessimismo quanto às possibilidades de realização do projeto moderno, com a falência dos mais caros ideais de emancipação social e a impossibilidade de suas ressuscitações.

Mas isso jamais poderia ter sido uma novidade para Habermas em 1986, uma vez que ele mesmo já havia confrontado teoricamente, com seu projeto de fôlego da razão comunicativa sistematizado em *Teoria da ação comunicativa*, de 1981, as teses inapeláveis da razão instrumental e da indústria cultural. Estudos robustos e sistemáticos sobre o abandono de perspectivas revolucionárias por parte dos antigos filósofos do Instituto, quase todos de alguma forma adeptos das teses de Benjamin sobre a história, já tinham sido publicados, e seus variados motivos eram mais do que conhecidos. Mesmo do outro lado do Atlântico, no Brasil, Olgária Matos já havia defendido sua tese sobre *Os arcanos do inteiramente outro* (apresentada como tese de doutoramento justamente em 1986 e publicada como livro em 1989) com um capítulo pioneiro dedicado à natureza pessimista da emancipação humana no último Horkheimer. Em seu *A fisionomia espiritual de Max Horkheimer*, publicado como estudo introdutório à edição póstuma de *Notizen*, de 1974 (e republicado como capítulo em 1977)[9], Alfred Schmidt sintetizara a "situação espiritual e histórico-política" que Habermas também conhecia bem: no ambiente ideal da Guerra Fria,

> contra o Ocidente [...] encontram-se alinhadas uma grande potência militar e uma agressiva ideologia stalinista. O marxismo, já brutalmente retomado nos anos do nacional-socialismo, continua a aparecer – aos olhos da expressiva maioria dos alemães-ocidentais, inclusive das organizações operárias – como doutrina inaceitável por definição, na medida em que é imediata-

[8] "A aporia com que defrontamos em nosso trabalho revela-se assim como o primeiro objeto a investigar: a autodestruição do esclarecimento. Não alimentamos dúvida nenhuma – e nisso reside nossa *petitio principii* – de que a liberdade na sociedade é inseparável do pensamento esclarecedor. Contudo, acreditamos ter reconhecido com a mesma clareza que o próprio conceito desse pensamento, tanto quanto as formas históricas concretas, as instituições da sociedade com as quais está entrelaçado, contém o germe para a regressão que hoje tem lugar por toda parte. Se o esclarecimento não acolhe dentro de si a reflexão sobre esse elemento regressivo, ele está selando seu próprio destino" (ADORNO; HORKHEIMER, 1985, p. 10).

[9] Na coletânea de A. Schmidt com três textos próprios, intitulada Drei Studien über Materialismus (München, Carl Hanser).

> mente equiparada ao regime soviético. Nestas circunstâncias, não há mais sentido para a escola recolhida em torno a Horkheimer insistir naquela unidade de análise crítica e práxis revolucionária, que se antes já aparecia como problemática, agora parece se fazer totalmente improvável (SCHMIDT, 1977, p. 104, trad. minha).

O capítulo de Habermas (*Observações sobre o desenvolvimento da obra de Max Horkheimer*, de 1986) não revelaria, pois, uma simples desatualização em relação a esse contexto histórico e teórico, não revelaria um inacreditável descompasso em relação ao "estado da arte", mas sim uma refutação proposital e pontual do referido texto de Alfred Schmidt (*A fisionomia espiritual de Max Horkheimer*, de 1974/1977), em especial sobre o estatuto filosófico do pensamento horkheimeriano tardio registrado em *Notizen* – refutação publicada numa coletânea organizada pelo próprio Alfred Schmidt –, inclusive com o uso de termos idênticos aos do interlocutor. Se Schmidt (1977, p. 109) afirma que as *Notizen* contém "a substância da última filosofia de Horkheimer", Habermas (1986, p. 276) afirma que "a substância da sua obra" reduz-se "aos trabalhos que apareceram antes do fim da Guerra". Se Schmidt considera que *Notizen* são "o testamento espiritual de Horkheimer", que "deveria substituir a grande obra da qual falava frequentemente em seus últimos anos, e que a ocasião do final o impediu de escrever" (SCHMIDT, 1977, p. 131), e que chegaria até mesmo a resultados semelhantes ao da *Dialética negativa* de Adorno (idem, p. 112), Habermas vê nelas um Horkheimer solitário, "que prolongou sua Teoria duplamente fraturada confiando-a às páginas de um diário" (HABERMAS, 1986, p. 293).

O que intérpretes como Post (1971), Schmidt (1977), Matos (1989) e outros não dispostos à reprovação sumária da produção tardia de Horkheimer delineiam sobre o contexto e os motivos da última Teoria Crítica horkheimeriana é inteiramente coincidente: o abandono do projeto original dos anos 1930 e as teses cada vez mais desencantadas da dialética do esclarecimento e da razão instrumental, acentuadas e reiteradas a partir de textos do exílio como *O fim da razão* (ou *Razão e autoconservação*) e *Eclipse da razão* (ou *Para a crítica da razão instrumental*), não acarretaram em deserdar a crítica e entregar-se à resignação na fase tardia. Apenas em reconhecer que, tanto para a pesquisa e a doutrina social quanto para o horizonte emancipatório proporcionado com auxílio delas, seria impossível não ter de ser mais modesto em relação às ambições das fases anteriores da teoria: "reduzem-se à tentativa [...] de compreender as profundas mudanças antropológicas em curso e de proteger os indivíduos dos pensamentos ou comportamentos ilusórios e sujeitos ao preconceito. Trata-se de uma contribuição à emancipação de seres humanos politicamente conscientes, os quais desejam um mundo mais justo e racionalmente fundado" (SCHMIDT, 1977, p. 116, trad. minha).

Para tanto, com a inserção cada vez mais explícita de Schopenhauer no quadro da Teoria Crítica, é o papel cada vez mais atuante de um pessimismo crítico e não-conformista que será elaborado; pessimismo, aliás, que ao contrário do que parece supor Habermas, não é mera decretação de desesperança. A esperança do pessimismo apenas é outra esperança que não a histórica, pois como elabora Matos (1993, p. 75), buscar a vitória histórica significaria "manter-se no registro do inimigo". Não por acaso, são desse período pós-retorno do exílio as frequentes conferências de Horkheimer sobre Schopenhauer e o pessimismo na sede da Schopenhauer-Gesellschaft, em Frankfurt[10]. Uma tentativa de síntese do que consistiria um pessimismo crítico ou pessimismo crítico social horkheimeriano desse período, em oposição ao pessimismo metafísico schope-

[10] Para um resgate dos contextos teóricos e dados biográficos e institucionais da maior aproximação de Horkheimer com a Sociedade Schopenhauer após seu retorno dos Estados Unidos, em especial sobre cada uma das conferências proferidas, cf. FAZIO, 2023b.

nhaueriano, em especial se notado a partir dos cinco artigos que dedicou a Schopenhauer de meados da década de 1950 até um ano antes de seu falecimento (ocorrido em 1973), seria a seguinte: i) a recusa terminante em atribuir genericamente a uma dinâmica própria do mundo volitivo as causas da falta de satisfação prometida e não cumprida pelo progresso; ii) um pessimismo que, em lugar de sobrecarregar "o mundo", filiaria os fracassos sociais a suas respectivas sociedades, dinâmicas e compromissos; isto é, a ideia de que um pessimismo metafísico em relação *à sociedade* precisaria ceder para um pessimismo crítico *das sociedades* ou de seres sociais específicos; iii) a crítica de uma cada vez menor importância do indivíduo e da cultura no interior da sociedade administrada, depois de Auschwitz, ou depois dos horrores de Hitler e Stalin. Neste caso, o diagnóstico de um pessimismo não resignado ou não conformista seria o de que "ao horror do passado sucederá um futuro administrado" (HORKHEIMER, 2024a, p. 230, trad. minha), com o que a humanidade se converteria em um gênero unitário como o de outros seres vivos; e que a fantasia, a religião, o anseio e o pensamento autônomo seriam ilusões superadas. Esse pessimismo, ao contrário daquele de Schopenhauer, seria incondicionado por não contar com qualquer alternativa quietista ou redentora de retorno à vontade universal daqueles que superam o egoísmo. O seu material seria abundante no próprio desenvolvimento da sociedade.

No entanto, tão importante quanto o papel acessório que esse pessimismo empresta para diagnósticos do presente no "mundo administrado" é a função que ele desempenha como mecanismo metodológico garantidor do *negativo* da crítica para suas proposições sobre o futuro, como instrumental de repulsa, ao mesmo tempo, ao positivismo e ao "otimismo conciliatório". Na referida entrevista em que conta a Gerhard Rein sobre como Pollock lhe apresentou Schopenhauer dizendo que "talvez você se interesse por isso [pelos *Aforismos para a sabedoria de vida*]", Horkheimer sintetiza, retrospectivamente e no ano anterior ao de sua morte, por que e em que termos "aquilo" lhe interessou muito. Rein havia perguntado a Horkheimer sobre a influência que sofrera de três pensadores: Schopenhauer, Freud e Marx. Na parte da resposta sobre Schopenhauer, o ilustre interlocutor afirma:

> Posso dizer que Schopenhauer não só desempenhou um papel importante na minha vida, mas inclusive a Teoria Crítica - como mais tarde foram chamados meus pensamentos - contém muito de Schopenhauer; pois ele diz, como eu, que não podemos descrever o que é absolutamente bom, mas diz: "Em última análise, a essência de todas as coisas é o mal, nomeadamente vontade de vida, a busca pelo bem-estar e pelo existir; e o verdadeiramente verdadeiro está no além - isso é o quão longe ele vai, mas não eu -, é o Nada". Então lidei muito intensamente com Schopenhauer (HORKHEIMER, 2024b, p. 451-452, trad. e grifos meus).

Se na edição de seus *Gesammelte Schriften* essa entrevista foi intitulada *Esperar o mal e ainda assim buscar o bem* (*Das Schlimme erwarten und doch das Gute versuchen*, cf. HORKHEIMER, 2024b), ela não deixaria de poder ser compreendida como inteiramente afinada à nossa hipótese, pois o bem a ser buscado aí já é suposto como conceito negativo, como negação do mal. Alfred Schmidt captou e sintetizou com clareza essa negatividade de toda e qualquer proposta emancipatória horkheimeriana – que, então, continua existindo: "Crítica – e isso é ainda mais raramente reconhecido – ela o é na medida em que, com Schopenhauer, afirma corajosamente que a espécie humana, inclusive quando pudesse ser mais sabiamente organizada, permanece ancorada em uma finitude radical: a espécie humana é algo de mísero no cosmos" (SCHMIDT, 1974, p. 139). É nesse horizonte que em um aforismo das *Notizen* com a data imprecisa de 1966-1969, intitulado justamente *Sobre a Teoria Crítica*, lemos uma explícita e renovada refutação de qualquer positivação da sua concepção tardia: "Ela [a Teoria Crítica] declara que o mal pode ser indicado – sobretudo na esfera social, mas também naquela do humano singular, na esfera moral

– mas *o bem, não*" (HORKHEIMER, 2008, p. 419, grifos meus). O que o pensador parece querer operar aqui, no limiar da década de 1970 e após tantas fases marcantes e de direções diversas que a Teoria Crítica originária das décadas de 1920 e 1930 já havia experimentado (em especial, pelos trabalhos dos outros "frankfurtianos"), é uma espécie de insistência renovada – que é, de forma proposital, apenas parcialmente uma novidade – em relação às teses centrais da *Dialética do Esclarecimento* e de *Eclipse da razão* (ou de *Para a crítica da razão instrumental*). Neste último texto, lemos o seguinte: "O que tem sido dito sobre a dignidade do ser humano é, por certo, aplicável aos conceitos de justiça e igualdade. Tais ideias devem preservar o elemento negativo, a negação do antigo estágio de injustiça ou iniquidade, e, ao mesmo tempo, conservar a significância absoluta original, enraizada em suas terríveis origens. De outra forma, elas tornam-se não apenas indiferentes, mas falsas" (HORKHEIMER, 2015b, p. 45).

Para a insistência do final da década de 1960, duas ilustrações são trazidas à baila: a proibição hebraica de representar Deus e a divisa kantiana de divagar em mundos inteligíveis contém ambas o reconhecimento da impossibilidade de determinar o Absoluto, ao que se segue a afirmação mais importante para a nossa questão: "o mesmo ocorre na Teoria Crítica" (HORKHEIMER, 2008, p. 420) na medida em que declara que o mal se refere substancialmente ao presente enquanto o bem precisa se confirmar como tal a cada vez, ou seja, não pode ter sua afirmação ou confirmação antecipada, pois isso transcenderia as possibilidades de quem julga e representaria a absolutização de uma hipótese. Daí, entre negativo e positivo, a conclusão só pode ser uma: "A análise crítica da sociedade denuncia a injustiça dominante; a tentativa de superá-la levou repetidamente a uma injustiça ainda maior" (idem, p. 420). Claro que Horkheimer não está advogando, aqui, furtar-se à ação ante o mal sob nossos olhos; de forma alguma defende a omissão diante dos males cotidianos, como um assassinato ou o escárnio da fome. Nenhum imobilismo é apregoado. Está se referindo à superação de tipos de injustiça e de opressão pela implementação de medidas que necessariamente conduzem a outras injustiças, e que o desenvolvimento e o progresso necessariamente produzem novos males: "A circunstância de que o desenvolvimento cego da tecnologia fortalece a opressão e a exploração social ameaça, a cada estágio, transformar o progresso em seu oposto, a completa barbárie" (HORKHEIMER, 2015b, p. 149). O determinante, nesse sentido, é a insistência de Horkheimer na importância do negativo: "Se se quer definir o bem como tentativa de abolir o mal, então é possível determiná-lo. Justamente esse é o ensinamento da Teoria Crítica. Já o contrário – ou seja, a tentativa de definir o mal por meio do bem – seria impossível, inclusive na moral" (HORKHEIMER, 2008, p. 420)[11].

Disso se colhe que a relação umbilical entre pessimismo e Teoria Crítica por meio do conceito de "negativo" poderia ser resumida nos seguintes termos: buscar o bem significaria negar o mal por meio da constante busca por "eliminar o pior" relativo a cada caso. Para garantir coerência semântica, um pessimismo filosófico (de teor schopenhaueriano) precisaria ser assumido, e empregado enquanto termo, como sinônimo e garantidor do "negativo" – dado ser impossível defini-lo sem a noção de mal como conceito positivo, o que acarreta a busca do bem apenas como negação do mal – e otimismo precisaria ser assumido como sinônimo de "positivo" ou de positivação – dado ser impossível defini-lo sem a noção de bem como conceito positivo, o que acarretaria a busca do bem como afirmação ou como absoluto. Essas premissas apoiam tanto o pensamento de Schopenhauer como influenciador de Horkheimer quanto, ao menos par-

[11] Cf. também HORKHEIMER (2008), aforismos intitulados *Dificuldades com o mal* e *O mal na história*.

cialmente, as sistematizações declaradas do pessimismo que se seguiram à obra do autor de *O mundo como vontade e representação*, com modificações da sua metafísica da vontade por Eduard von Hartmann e pelos outros pensadores e pensadoras do *Pessimismusstreit* da segunda metade do século XIX. Horkheimer, se não foi influenciado diretamente pelos autores desse grupo[12], o foi confessamente por Schopenhauer como fonte comum. Como motivo basilar dessa influência, está todo o conteúdo negativo do pensamento schopenhaueriano – que Horkheimer soube captar de forma sutil –, espraiado em uma variedade de conceitos: o bem, a liberdade, o prazer e a felicidade apenas negam a primazia do mal, a determinação ou a opressão, a dor ou os sofrimentos.

Frente a Schopenhauer como sua fonte para o mal positivo metafísico, a tarefa que Horkheimer se coloca direta ou indiretamente pode ser registrada nos seguintes termos: a partir do consenso quanto a premissas básicas opostas às de todo otimismo filosófico, como aquela da positividade do mal e do sofrimento em detrimento da negatividade do bem e da felicidade, e a da preferência pelo não-ser em lugar do ser, que tipo de ação supra-individual agregaria sentido para ser socialmente motivada, defendida e promovida? Ou seja, como mobilizar socialmente o pessimismo a favor da crítica emancipatória? Se o nominalismo de Schopenhauer em relação à sociedade e à concessão de "realidade" apenas aos indivíduos não o deixaram propor saídas sociais e políticas robustas, a sua reiterada não-justificação do existente e seus esboços de crítica social, em especial da obra tardia[13], parecem ter bastado para que parte de seus intérpretes assumisse a elaboração de projetos socialmente engajados, ou diretrizes de projetos, como decorrência do pessimismo metafísico. Horkheimer, que assim contradiria a avaliação de Habermas, não deixou de ser um deles, em especial na fase tardia. O rebaixamento do fator histórico-social ou de alguma totalidade não impediria a práxis em geral. Impediria apenas a práxis de tipo marxista, inclusive aquela não-ortodoxa nos moldes propostos nos anos 1930. A práxis seria reconfigurada com fatores menos pretenciosos, e certamente menos utópicos.

Claro está que, para isso, se considera superada a limitação da ideia de que pessimismo seja necessariamente sinônimo de quietismo, resignação ou imobilismo, algo que só quem insiste em um pessimismo monossêmico continuaria negando. Se minimamente "crítico", mesmo se não em sentido estrito do adjetivo "crítico" de uma Teoria *Crítica*, não podemos mais compreender pessimismo meramente como versão filosófica do seu emprego popular, usado, em geral, para expressar desânimo, falta de esperança ou expectativa ruim em relação ao futuro. A chamada "esquerda schopenhaueriana", que segundo Lütkehaus (2007) tem o próprio Horkheimer como um dos seus fundamentadores iniciais, já foi suficientemente definida[14] e propicia atualmente novas elaborações e propostas.

[12] Na obra publicada e também no espólio geral disponível nos Arquivos Horkheimer da Goethe-Universität Frankfurt não se encontram citações, referências ou indicações diretas de Horkheimer sobre autores do *Pessimismusstreit* da segunda metade do século XIX, como a Eduard von Hartmann, Agnes Taubert, Olga Plümacher, Julius Bahnsen e Philipp Mainländer (cf. HORKHEIMER, 2024c).

[13] Horkheimer cita de forma reiterada as denúncias de Schopenhauer contra a escravidão negra e a exploração do trabalho, em especial do trabalho infantil, nas fábricas de sua época. As críticas são elaboradas pelo filósofo da vontade no Tomo II de *O mundo como vontade e representação* (cf. SCHOPENHAUER, 2015, Cap. 46) e no Tomo II de *Parerga e paralipomena* (cf. SCHOPENHAUER, 2012, § 114 e 125).

[14] A fundamentação da chamada "esquerda schopenhaueriana" conta, hoje, com pelo menos os seguintes principais sustentadores (em ordem cronológica): LÜTKEHAUS, 1985, 2006, 2007; DEBONA, 2013, 2020, 2022; DURANTE, 2018, 2022; CACCIOLA, 2022; CIRACÌ, 2022; FAZIO, 2023a.

III

Mas se é assim, e se foi de Schopenhauer que Horkheimer assimilou o mal como conceito positivo para as tarefas da crítica da sociedade, o bem como conceito negativo - porque existe apenas na medida em que *nega* o mal -, e, por conseguinte, o conceito de pessimismo como pessimismo *crítico* e como expressão do negativo, como, então, faria sentido e seria coerente a orientação reiterada que encontramos em outros escritos tardios sobre o pessimismo *teórico* precisar corresponder a um otimismo *prático*, ou a uma práxis otimista? Seria relevante indicar em que medida essa espécie de oscilação em relação ao papel consequente de um pessimismo na formulação da Teoria Crítica originária horkheimeriana pode confundir e prejudicar. Ela pode ter afetado a compreensão de Habermas? A nosso juízo, se teria potencial para tanto, uma consideração sobre o emprego impreciso dos termos "pessimismo" e "otimismo" e seus sentidos não compromete o essencial, isto é, o conteúdo material - filosófico e sócio-político - do que pode ser chamado de *esboço de emancipação anti-otimista* na última Teoria Crítica de Horkheimer. Vejamos.

Uma das ocasiões em que elaborou a referida divisa de pessimismo teórico e otimismo prático foi no final do texto Teoria Crítica ontem e hoje (1969), escrito algumas semanas após o falecimento repentino de seu amigo Adorno, e que foi apresentado em pelo menos duas conferências antes de ser publicado em 1970:

> Para concluir, gostaria de dedicar uma palavra sobre a diferença entre pessimismo e otimismo. A minha concepção de culpa do gênero humano é efetivamente pessimista; e pessimista é a convicção de que a história caminha em direção a um mundo administrado, com o que aquilo que chamamos de espírito e fantasia acabará por regredir em grande medida [...]. Mas, então, em que consiste o otimismo que partilho com Adorno, o meu amigo falecido? Na convicção de que, apesar de tudo, se deve buscar e tentar realizar aquilo que se retém como verdadeiro e bom. *Era esse o nosso princípio: pessimista na teoria e otimista na prática* (HORKHEIMER, 2022, p. 353, trad. e grifos meus).

Qual pessimismo para a teoria e qual otimismo para a prática? E como poderia uma práxis negativa, com aquela carga conceitual mais schopenhaueriana do que marxista, reivindicar como seu aliado um *otimismo* prático? A reincidência da ideia pode nos esclarecer alguns pontos. Ela acontece no artigo *Pessimismo hoje*, escrito naquele mesmo ano de 1969 para ser apresentado como conferência na Sociedade Schopenhauer, e publicado em 1971 no Schopenhauer-Jahrbuch: após registrar novamente seu diagnóstico de um mundo completamente administrado como resultado do progresso avassalador da razão instrumental em detrimento da autonomia individual, dos valores culturais e do espírito, Horkheimer afirmará que a quem é consciente "da miséria do passado, da injustiça do presente e da perspectiva de um futuro sem significado espiritual" (HORKHEIMER, 2024a, p. 232, trad. minha) restaria apenas o anseio (*Sehnsucht*) constantemente ameaçado pelo progresso; mas que, *se essas pessoas resistentes se encontrassem*, não seria insignificante o que ainda poderiam realizar para "aliviar os sofrimentos humanos" (ibidem): poderiam consentir uma forma de solidariedade que compreenderia *aspectos* teológicos não dogmáticos, sendo que (o que nos interessa em particular) suas atitudes, "em última análise *negativas* (*negative Haltung*), seriam consoantes àquela que aqui em Frankfurt é chamada 'Teoria Crítica'" (ibidem). Essa práxis negativa resultante de pessimismo e Teoria Crítica é assim elaborada:

> Os seres humanos unidos por esse anseio não poderiam afirmar nada sobre o Absoluto, sobre uma realidade puramente inteligível, sobre Deus e a redenção, não poderiam atribuir um valor de verdade absoluta ao seu conhecimento, a nenhuma forma de conhecimento; poderiam, no entanto, difundir a solidariedade, indicar – levando em conta aquele progresso que é necessário, embora deva ser pago com um alto preço – o

> que deve ser mudado ou preservado para aliviar os sofrimentos humanos (HORKHEIMER, 2024a, p. 232, trad. minha).

Ora, aqui se encontra um elemento central para aquilo que podemos compreender como esboço de projeto emancipatório negativo ou anti-otimista da fase tardia horkheimeriana: admitindo-se formas de organização social, algo como "comunidades de resistências" pautadas pelas mais diversas causas poderiam se formar em torno de uma ideia pessimista de solidariedade, essa que por si mesma já é necessariamente apenas a negação de determinada opressão na imanência do mal social. Ao contrário da afirmação de Habermas de que Horkheimer teria reduzido toda e qualquer saída ao terreno da teologia, está claro que esse tipo de práxis, se admite "aspectos" ou "momentos" teológicos, não se reduz a eles. O problema, então, não seria exatamente aquele apontado por Habermas, mas sim o que surge do fato de essa espécie de chamamento pessimista de Horkheimer à ação ser finalizada com os seguintes termos: "Ao *pessimismo teórico* poderia ser associada uma *práxis não anti-otimista* (*nicht unoptimistische Praxis*) – que, ciente do mal universal, tentasse melhorar o mundo tanto quanto possível" (HORKHEIMER, 2024a, p. 232, trad. minha).

Ou seja, uma práxis, em última instância, otimista. Se a dupla negação da língua alemã (*nicht* + *un*) poderia denotar um contorcionismo retórico de Horkheimer para não advogar expressamente uma *práxis otimista* ou um *otimismo prático* em um artigo intitulado *Pessimismus heute*, publicado no Schopenhauer-Jahrbuch de 1971 após ser apresentado como conferência em evento da Sociedade Schopenhauer cerca de dois anos antes (em novembro de 1969), essa indicação expressa foi registrada no já citado *Teoria Crítica ontem e hoje*. Não é mera coincidência o fato de se tratar do mesmo texto em que reiterou terem sido Schopenhauer e Marx os dois filósofos que mais influenciaram os inícios da Teoria Crítica. A compreensão dessa aparente estranheza dependeria em grande medida do teor da práxis em questão. Nessa época, sabemos ao menos que ela está despida, há muito tempo, de sua antiga roupagem em tons predominantemente marxistas, da ideia de uma "sucessão histórico-universal de fases materialisticamente concebidas" (SCHMIDT, 1977, p. 114). Trata-se mais, para esse "último Horkheimer", de uma espécie de "práxis schopenhaueriana" que não abandona tudo o que pode germinar do marxismo não-ortodoxo das fases anteriores. E, ainda assim, seria possível compreender que Horkheimer tenha dissociado teoria e prática do pessimismo para, ao mesmo tempo, se distanciar de Schopenhauer, para quem, como lemos no Tomo II de sua obra principal, "se expresso PRATICAMENTE, [o mundo] não deveria ser, TEORICAMENTE ele também não seria um problema" (SCHOPENHAUER, 2015, p. 662), o que implicaria para Horkheimer definir também a práxis do pessimismo teórico como questão metafísica de ser ou não ser, ou da preferência pelo não-ser. Isso realmente não poderia definir um pessimismo crítico ao modo da Teoria Crítica, inclusive porque Horkheimer não reconhece todas as conclusões da metafísica schopenhaueriana.

Ainda sobre essa espinhosa questão, há mais um elemento. Ele pode ser colhido das mesmas *Notizen*, em especial de um fragmento de 1961-1962, em que lemos uma curiosa crítica pontual ao que Horkheimer chama de "o otimismo de Schopenhauer". A crítica é dirigida fundamentalmente às teses da negação da vontade expostas no Livro IV de *O mundo como vontade e representação*. O frankfurtiano acusa o "Buda de Frankfurt" de ser contraditório em relação a seu pessimismo devido às teses de metafísica imanente que admitem a soteriologia da vontade individual, mediante quietismo espontâneo como primeiro passo para o retorno da vontade cindida à vontade una: "Seja lá o que for que um ser humano sonhe como final do sofrimento, morte e

ressurreição, o que afirme de forma absoluta, o amor celeste ou terrestre: tudo isso não é mais que um instante de falsa infinitude. A boa infinitude é um consolo duvidosamente filosófico. Dessa maneira, em última instância, Schopenhauer se confirma contra si mesmo" (HORKHEIMER, 2008, p. 388, trad. minha). Mesmo sabedor da natureza não-ensinável e extremamente rara do ascetismo schopenhaueriano, Horkheimer se incomoda com o otimismo redentor deixado como única e remota brecha para libertação dos horrores terrenos. A tese "demasiado afirmativa" de Schopenhauer, como elabora Schmidt, pareceu a Horkheimer se sobrepor à verdade de que "a dor é eterna".

Mas se, como indicado acima, Horkheimer se tornou defensor de um tipo de otimismo em sua fase tardia – em seu caso, ao contrário de Schopenhauer, expressamente admitido –, ele não teria incorrido no mesmo problema que acusara em Schopenhauer, ainda que com outro conteúdo? Afinal, o conteúdo filosófico do que chamou de "otimismo prático" ele mesmo teria fornecido, como está evidente na elaboração de 1971, quando menciona expressamente o "anseio" (*Sehnsucht*): se trataria, em suma, de tudo aquilo que compõe a sua defesa do "anseio pelo inteiramente Outro", incluindo para tanto o papel de uma teologia negativa e uma clara esperança: "[...] A esperança de que essa injustiça que caracteriza o mundo não permaneça, de que a injustiça não seja a última palavra. [...] Um anseio de que o assassino não triunfe sobre a vítima inocente" (HORKHEIMER, 2024b, p. 389, trad. minha). Isso, portanto, levaria o leitor a entender que estivesse defendendo uma declarada e proposital positivação (cf. RAMOS, 2017), mesmo com a reiterada afirmação de que a teologia não dogmática e a solidariedade reivindicadas para tanto seriam necessariamente *negativas*; pois, com isso, se estaria buscando de qualquer modo um mundo justo, ao invés de um menos injusto.

IV

Porém, um suposto "Horkheimer contra Horkheimer" deixa-se entrever como improcedente nos fragmentos de *Notizen*. Apesar das referidas indicações em vista de uma "práxis otimista" como a face prática do pessimismo teórico, há uma paralela suposição de que nenhum otimismo – filosoficamente justificado – sustentaria o tipo de práxis anti-utópica e anti-finalista que ainda restaria. A aparente contradição se desfaz quando notamos que o emprego de "pessimismo" ou "pessimista" e de "otimismo" ou "otimista" nem sempre é feito em sentido estritamente filosófico-crítico no sentido da tradição schopenhaueriana, e que, nas ocasiões em que registrou o "otimismo prático" como necessário, Horkheimer não se referia rigorosamente ao conceito de otimismo da tradição filosófica moderna, recusado como falso e perverso por Schopenhauer e por ele mesmo por justificar as dores e os males em nome do bem ou do progresso. O otimismo filosófico, conceitualmente considerado e alvo da crítica de tradição schopenhaueriana, opera ao lado das "reconciliações idealistas" em relação ao *status quo*; e a ele a Teoria Crítica agregou a resignação, retirando-a do pessimismo: "A Teoria Crítica recusa qualquer otimismo que confira à *objetividade* – entendida seja como *progresso*, seja como *finalidade* – a 'realização da história' [...]; recusa a espera otimista de que do próprio curso do mundo haverá a vitória do sentido e da razão" (MATOS, 1989, p. 254). Desse modo, o que Horkheimer chama de "otimismo prático" ou de "práxis não anti-otimista" não poderia, apesar da terminologia, ser compreendido como portador de semântica correspondente a seu par "pessimismo teórico", ou seja, de um sentido oriundo da tradição que ele herda e reelabora enquanto pessimismo crítico, componente basilar do que Lütkehaus (2007) dissera se tratar de uma "esquerda schopenhaueriana".

Em vez disso, se trataria mais de um otimismo em sentido cultural e de uso coloquial. Essa compreensão salvaria a coerência em relação à continuidade de uma sintonia parcial com a ideia marxiana de práxis – que não dissocia teoria de prática –, mesmo se considerarmos que essa fase tardia de Horkheimer é justamente menos fiel a Marx. É como se Horkheimer se permitisse uma dinâmica conceitual menos rígida em relação a seus principais influenciadores dos inícios da Teoria Crítica (Marx e Schopenhauer), uma hipótese que pode ser atestada por outro fragmento das *Notizen* intitulado *Otimismo socialmente necessário*: "Que os seres humanos sejam alegres e de bom humor, que digam sim à vida, exprime a finalidade à qual a cultura se propõe, por mais horrível que seja o fundamento do mundo, o modo com que é organizado, a cadeia funesta da história, a morte na dor, na angústia, na miséria" (HORKHEIMER, 2008, p. 390, trad. minha). O fato de "o otimismo dos governados" ser determinante para o êxito de governos não se confunde com a defesa de um otimismo filosófico como doutrina justificadora do mundo e seus males, em que "a *mentalidade afirmativa* com a qual o horror da realidade não se torna superado, contribui apenas para perpetuá-lo" (ibidem, grifos meus). Ou seja, não haveria uma contradição em termos. O pessimismo crítico-filosófico e consequente da última Teoria Crítica horkheimeriana estaria garantido e poderíamos compreender suas hipóteses a partir de um anti-otimismo, não obstante a recomendação de um tipo popular e cultural de otimismo para a ação.

Por essa senda, podem ser reconhecidos alguns dos conteúdos mais específicos do que sugiro chamar de esboços de emancipação anti-otimista da obra tardia de Horkheimer. Estão espalhados pela referida coletânea de apontamentos (*Notizen*), que também poderiam ser considerados aforismos de política não-radical. "Anti-otimista" e "anti-positivista" teriam de poder ser considerados, nesses esboços de direcionamentos, como sinônimos, ao passo em que expressam o que significa o acima mencionado horizonte emancipatório "mais modesto" em relação às pretensões da primeira Teoria Crítica horkheimeriana. Os esboços anti-utópicos, de "política negativa" e de anti-extremismo, performariam elementos para projetos emancipatórios *anti-otimistas* no plano crítico-filosófico *stricto sensu*, ainda se otimistas no plano cultural *lato sensu*.

Um norte pode ser localizado no aforismo intitulado *Negativos*, em que lemos: "Os espíritos negativos, negativistas, que veem e dizem apenas o que é horrível, apenas o que não deve ser, que têm medo de nominar Deus, o que esses espíritos, afinal, desejam? Que as coisas melhorem. Os positivistas agem em Seu nome, dizem sim ao mundo e ao Criador. Unem-se – não são contra os sacros valores. Os têm sempre na ponta da língua. Assim Hitler uniu os alemães, fazendo dos judeus a vítima designada; Nasser os árabes, designando Israel ao papel de vítima" (HORKHEIMER, 2008, p. 240). Não é necessário considerar os positivistas que agem em nome do bem chamado Deus; seriam suficientes os que o fazem em nome de qualquer bem. Esse norte pode também ser captado no aforismo intitulado simplesmente *Teoria Crítica*, em que Horkheimer ironiza o fato de que se continue exigindo incessantemente da Filosofia indicações práticas sobre "o que precisa ser feito", indicando que a Filosofia "não encontrou nenhum novo céu para poder apontar, nem mesmo um céu terreno; e que a sua descoberta é justamente a de que "o céu ao qual se pode indicar um caminho não é um céu" (idem, p. 253). E, ainda, no aforismo *Os três erros de Marx*, que denuncia a concepção burguesa de sociedade e de liberdade pela denúncia dos horrores e da miséria que precisariam ser aceitos em nome da busca positiva por essa liberdade; bem como a crença de Marx de que a paz a ser alcançada entre as classes significaria também a paz entre os homens e em relação à natureza.

Por isso mesmo, a emancipação anti-otimista aqui indicada precisaria mais de Schopenhauer

do que de Marx: "O materialismo de Marx, quando livre do autoengano idealista, se aproxima mais de Schopenhauer do que de Demócrito" (HORKHEIMER, 2008, p. 270, trad. minha). No reino da liberdade, Horkheimer não reconheceria mais do que "a solidariedade com a vida, a luta pela justiça não apenas na sociedade, mas na natureza em geral". Essa solidariedade que, como formulou magistralmente Olgária Matos (1989, p. 253), "nós a conquistamos graças à desesperança". Por essa via, é Schopenhauer quem fomenta a práxis porque fica do lado do que é temporal e não do que é desapiedadamente eterno; a solidariedade como conceito negativo e suas potencialidades no âmbito de uma política negativa fundamenta-se na ideia de que a união dos humanos, ação que nega e resiste, deve-se ao desconsolo e ao desamparo como signos da positividade do mal.

Mas é no aforismo intitulado *Política negativa* que encontramos expressamente um exemplo prático fundamental do que estou propondo chamar de emancipação anti-otimista – para não dizer emancipação negativa. Horkheimer começa observando que "também na *política* vale a *teologia* negativa" (HORKHEIMER, 2008, p. 260, trad. minha). Isto é, a ideia de que as conhecidas teses do "anseio pelo inteiramente Outro", defendidas como tendo, inclusive, momentos ou aspectos teológicos, teriam o mesmo ou ainda maior espaço na política. A hipótese basilar, que intitula o presente artigo e com a qual o pensador conclui o aforismo, é registrada por Horkheimer como uma espécie de lema: "*Eliminar o pior é mais humano do que buscar o bem*" (idem, p. 261). O problema posto corresponde, em outros termos, à pergunta sobre em que medida seria possível defender um projeto libertário ou emancipatório sem o apelo a um futuro *melhor* – utópico ou não – em sentido estrito e positivo, dado que "melhorar" dificilmente não é associado a um *bem* (positivo) que precisaria ser buscado como tal. O problema não seria meramente retórico com os qualificativos *melhor* e *pior*. Bastaria, para evitar essa dificuldade linguística inerente à busca por um mundo melhor e seu bem correspondente, substituir a positividade do melhor por uma linguagem que garanta algo de negativo? Buscar um mundo *menos pior*?

O caso escolhido pelo pensador para ilustrar essa direção de sua última Teoria Crítica soa de extrema atualidade: "Os programas de planificação econômica, que hoje se reduzem quase que inteiramente à apologia do *trend* frequentemente verificável nos Estados Unidos, têm a sua razão de ser em uma exigência de justiça. Mas não se sabe se a miséria pior – aquela que pede socorro nos cárceres e nos manicômios – se deva mais à intervenção brutal do que à planificação econômica" (idem, p. 260-261). A ideia, atualizadora da tese acima citada de *Eclipse da razão*, é a de que a justiça exigida estaria, no final das contas, identificada ao sistema injusto que, em nome de tendências econômicas – hoje, diríamos tendências de mercado – tornaria as realidades sociais ainda mais cruéis, com maior alcance de seus efeitos, ou apenas substituiria um mal por outro. Se trataria de uma "antinomia da Teoria Crítica", conforme expressa o título de outro aforismo, que "reconhece que a injustiça é idêntica à barbárie, mas que a justiça é inseparável daquele processo tecnológico que transforma a humanidade em refinada espécie animal que reduz o espírito a uma manifestação superada da própria infância" (idem, p. 423). A aporia é repetida na entrevista sobre "o anseio pelo inteiramente Outro", de 1970: "Justiça e liberdade são conceitos dialéticos. Quanto mais justiça, menos liberdade; quanto mais liberdade, menos justiça. Liberdade, igualdade, fraternidade é um slogan maravilhoso. Mas se se quer manter a igualdade, então se tem que restringir a liberdade, e se se quer deixar as pessoas livres, então não pode haver igualdade" (HORKHEIMER, 2024b, p. 403, trad. minha). Uma atualização em termos de "mundo administrado" da antiga aporia da autodestruição do esclarecimento da *Dialética do*

Esclarecimento.

Essa antinomia da Teoria Crítica é a mesma do próprio pessimismo crítico que não cede seu papel, pois em outro aforismo de *Notizen*, intitulado justamente *Sobre o pessimismo*, é resumida a ideia acima referida da substituição de um mal por outro: a lógica imanente do máximo desenvolvimento social e, com isso, de supostas superações de mazelas do passado e do presente, não consegue evitar que, ao final, a vida se torne completamente automatizada: "O domínio do ser humano sobre a natureza consegue uma extensão tal que chega a comportar nela o desaparecimento da penúria [...], mas, ao fim e ao cabo, é também total desencanto, extinção do espírito" (HORKHEIMER, 2008, p. 420, trad. minha). Daí, mais uma vez, não obstante a certeza pessimística de que o saldo não será positivo e o mundo continuará *pessimus*, não deriva qualquer resignação, mas a clareza de que "hoje, a Teoria Crítica precisa [...] se referir ao chamado progresso, ou seja, ao progresso técnico, e aos efeitos que ele produz nos seres humanos e na sociedade; [...] denunciar a dissolução do espírito e da alma, a vitória da racionalidade instrumental, sem se limitar a refutar essa vitória" (idem, p. 423).

Outro elemento desses esboços emancipatórios pode ser identificado a partir de uma conjugação do aforismo intitulado *Contra o radicalismo de esquerda*, com o último aforismo de *Notizen*, *Para o não-conformismo*: a indicação é a de que "um elemento teórico do não-conformismo poderia ser, hoje, a análise crítica dos demagogos, assim como o associar-se de pessoas que o desenvolvam nos planos psicológico, sociológico e tecnológico poderia representar um seu momento prático" (idem, p. 425). Ora, não seria esse um claro projeto horkheimeriano de sua última Teoria Crítica, parcialmente já em execução por ele, com novas pesquisas empíricas após seu retorno do exílio, e, como se sabe, por Adorno e seus colaboradores nos EUA, sobre a personalidade autoritária? Que seja apenas um esboço de projeto talvez já baste para negar a inércia e o vácuo acusados. E que se trate de um projeto renovado, que se apresenta de alguma forma conservador, parece inegável, pois junto à defesa de que não faria sentido atacar o capitalismo sem a preocupação quanto ao perigo de totalitarismos, sem a compreensão da tendência ao fascismo no interior dos Estados capitalistas, mas também dos riscos da queda da esquerda radical em totalitarismos terroristas, tudo isso não deveria prescindir – e, ao contrário, agregaria coerência – em relação à ideia de que "uma resistência séria contra a injustiça social compreende necessariamente a defesa daquelas formas de liberdade de ordem burguesa que não deveriam desaparecer, mas, ao contrário, serem estendidas a todos os indivíduos" (idem, p. 414). A forma burguesa de liberdade, ao passo em que é estendida a todos, não seria mais forma burguesa de liberdade. O anti-radicalismo político de Horkheimer permite reivindicar, então, as liberdades e os direitos de alguns para toda a sociedade. Isso não representaria qualquer empecilho ou impedimento para movimentos e associações de base se organizarem em vista da preservação de autenticidades culturais. Não haveria conflito entre a ideia de uma comunidade pessimista de resistentes e a sua reivindicação de que um mundo "menos pior" pode ser democratizado. Pelo contrário, essa poderia ser uma de suas tarefas primárias. Apesar da incomensurável capacidade destrutiva do mundo administrado que, por exemplo, ameaça cada vez mais individualidades e singularidades, essa mesma singularidade "pode intervir criticamente no processo, tanto no plano teórico quanto no prático, contribuindo com métodos atuais para a formação de coletivos extemporâneos" (idem, p. 424-425).

Daí que à recomendação horkheimeriana de "pessimismo ao passado, otimismo ao futuro",

implícita em seus textos tardios sobre a definição de Teoria Crítica, caberia melhor a que indicasse "pessimismo ao passado e ao presente, anti-otimismo ao futuro". Lutas emancipatórias, de ontem ou de hoje, não se beneficiariam tanto de pessimismos, críticos ou menos críticos, se esses são entendidos como expressão de desesperanças generalizadas. Mas uma práxis calcada na consciência histórica de que a busca pelo bem positivo necessariamente implica em apenas alternar os papéis entre dominados e dominadores teria de ser anti-otimista se o par pessimismo-otimismo é assumido nos termos filosóficos da tradição crítica schopenhaueriana.

Se esses potenciais podem ser identificados nos textos tardios de Horkheimer, em especial nas *Notizen*, Habermas teria pouca razão ao declarar a opacidade emancipatória dos últimos anos do *spiritus rector* da Escola de Frankfurt. Se há questões problemáticas na última fase do pensamento de Horkheimer, em suas últimas elaborações do que entendia por "Teoria Crítica" e por seus papeis, o problema não consistiria no suposto conformismo acusado, ou em um suposto "pessimismo conformista", ou ainda em um tipo de "schopenhauerianismo resignado", mas sim justamente na falta de uma maior insistência e clareza por parte de Horkheimer quanto a um pessimismo não-conformista. Uma das amostras dessa falta encontra-se na recomendação analisada acima de que a um pessimismo teórico deveria corresponder uma "práxis otimista", o que nos leva a questionar o próprio sentido do emprego do termo "pessimismo", que talvez não possa ser assumido como tão schopenhaueriano quanto parece, ou como exclusivamente no sentido da tradição filosófica schopenhaueriana. No limite, não será desarrazoada a compreensão de que o Horkheimer tardio, por exigências postas por ele mesmo, em especial quanto ao conceito de "negativo", teria de ter sido mais schopenhaueriano – e não o contrário.

Os esboços de projetos emancipatórios anti-otimistas que identificamos em *Notizen* e nos outros textos supra analisados da mesma época têm como norte – e talvez por isso tenham permanecido apenas esboçados – o que Horkheimer responde a Claus Grossner na entrevista de 1971, editada como *Para o futuro da Teoria Crítica*. O entrevistador pergunta a Horkheimer sobre se (e como) ele se sentiria afetado por uma declaração de Habermas dada alguns dias após a morte de Adorno, em que afirmava ter sido aquele acontecimento a queda do último véu que encobria a "nudez metodológica" de que sofria a Teoria Crítica nos últimos tempos; e, mais ainda, que se, segundo Habermas, a intenção da Teoria Crítica permanece, "sua comunicação com análises sociais concretas, a integração de resultados empíricos científicos específicos em uma análise da sociedade como um todo está se tornando cada vez mais difícil" (HORKHEIMER, 2024b, p. 419, trad. minha). Horkheimer praticamente ignora o primeiro ponto da pergunta, limitando-se a dizer que "precisamos desenvolver suas ideias [de Adorno]", mas responde o seguinte sobre o segundo ponto: "A relação entre o projeto conceitual e o material preparado por cientistas individuais deve ser determinada novamente em cada caso. [...] A Teoria Crítica não consiste em nada mais do que no princípio de abster-se de apresentar a sociedade correta (*richtige Gesellschaft*), o bem absoluto, em termos de conteúdo, mas sim em criticar a sociedade atual, ou seja, identificar claramente o que pode e precisa ser mudado. O bem absoluto não está contido como algo positivo na própria teoria" (ibidem). Ou seja, o norte para projetos emancipatórios do último Horkheimer consiste na insistência de que, quaisquer que sejam os projetos, que então podem ser os mais diversos, teriam necessariamente de ser negativos e particularizados. Negativos quanto ao objeto em questão, no sentido de não apresentarem positivamente alguma solução para além da crítica, em vista de não correrem o risco de substituir um mal por outro; particula-

rizados no sentido de não se pretenderem projetos "para o todo"[15]. Em uma palavra, "eliminar o pior" referente a cada mal, em vez de "buscar o bem".

Habermas, em seu texto de 1986, tem razão apenas na afirmação de que Horkheimer não elaborou exatamente um projeto emancipatório *novo* em sua última Teoria Crítica. O que temos na última produção horkheimeriana, em especial nas *Notizen*, é mais uma continuação matizada e atualizada de teses da *Dialética do Esclarecimento* e de textos individuais posteriores (como *Eclipse da razão* ou *Para a crítica da razão instrumental*), com desenvolvimentos do mesmo tipo de preocupação, embora não exatamente a preservação de todos os elementos teóricos. Nela há também ênfases novas em alguns desses elementos a partir de diagnósticos parcialmente renovados. É o que confirma um dos últimos fragmentos de *Notizen*, o já citado *Sobre o pessimismo*: "Tudo isso retorna à dialética do esclarecimento, para a qual a verdade se arruína na adaptação incondicionada ao absurdo, à realidade enquanto tal" (HORKHEIMER, 2008, p. 420, trad. minha). Improcedente, porém, é a crítica de Habermas sobre as "desesperanças" horkheimerianas, que a seu ver seriam, sem mais, atestados de resignação. Horkheimer elaborou, como procurei mostrar e ainda que apenas como esboço, uma forma específica de garantia da negatividade para projetos emancipatórios. A filosofia de seus últimos anos é, no mínimo, um projeto de resistência em relação ao mundo da total-administração. Anti-otimista – ou criticamente pessimista – ele o é na medida em que é anti-positivo. Seu conceito de negativo, se não equivale ao da dialética negativa de Adorno, existe e foi formulado com tintas schopenhauerianas, o que Habermas não reconhece. Hoje, quando alguns habermasianos admitem que o grande projeto da razão comunicativa se vê eclipsado (CORTINA, 2024) na sociedade da Inteligência Artificial e das Big Techs, talvez os esboços de emancipação modesta e anti-otimista de Horkheimer façam mais diferença do que um grandioso projeto afirmativo.

[15] Türcke analisou de forma lúcida e certeira o horizonte emancipatório do Horkheimer tardio: "A filosofia tardia de Horkheimer não desabrocha. Todavia, a inconstância de seus Apontamentos [*Notizen*] revela, de uma forma decididamente existencial, um profundo impulso de Teoria Crítica: quanto mais ela quer expressar o que é verdade, tanto menos ela quer acabar por ter razão ao final. Relativamente a isso, ela é herança dos profetas do Antigo Testamento, que anunciavam a desgraça de maneira tão apodítica, para que, por fim, ela não acontecesse. E, assim como os profetas eram arrastados de um lado para o outro pelo paradoxo de sua tarefa, do mesmo modo também ocorre com a Teoria Crítica. Cada objeto de sua crítica é também uma tentação para ela. Todos lhe murmuram: Desista disso! A vida seria, afinal, infinitamente mais fácil se existisse um sentido superior ou mais profundo, no qual se pudesse confiar e do qual se pudessem extrair diretivas seguras. Também seria mais fácil se existisse uma grande esperança como o proletariado, um *black power* ou *women's power*, com o talento de arrastar consigo toda a humanidade e movê-la para um estado superior. E como seria belo se livrar sem dificuldade do invólucro de aço da era moderna, como a era pós-moderna tem em mente, ou a virada linguística realmente poderia favorecer o entendimento interpessoal dessa maneira e humanizar o capitalismo até torná-lo irreconhecível, como espera a teoria do agir comunicativo. E aquele que nunca foi apoderado pela tentação de evitar a miséria existente por meio de uma ou de outra das formas mencionadas, ou já fez as pazes com ela ou então não é deste mundo" (TÜRCKE, 2019, p. 180).

Referências

ADORNO, T. W. 2009. *Dialética negativa.* Trad. Marco Antonio Casanova. Rio de Janeiro: Zahar.

CACCIOLA, M. L. 2021. Nota à tradução. In: SCHMIDT, A. *Schopenhauer e o materialismo*. Trad. Maria Lúcia Cacciola. São Paulo: Clandestina.

CACCIOLA, M. L. 2022. Che cosa significa una lettura di sinistra del pensiero di Schopenhauer. In: D. FAZIO (Ed.); M. VITALE (Ed.). *Prospettive. Tredici saggi a duecento anni dal Mondo come volontà e rappresentazione di Arthur Schopenhauer*. Lecce: Pensa MultiMedia.

CHIARELLO, M. 2001. *Das lágrimas das coisas:* estudo sobre o conceito de natureza em Max Horkheimer. Campinas: Editora da UNICAMP; São Paulo: FAPESP.

CORBANEZI, E. 2017. Schopenhauer entre Marx e Schopenhauer: do materialismo pessimista ao pessimismo materialista. *Transformação*, Marília, v. 40, n. 4, p. 111-132.

CORTINA, A. 2024. *Ética o Ideología de la Inteligencia Artificial?*: El Eclipse de la Razón Comunicativa em uma Sociedade Tecnologizada. Barcelona: Paidós.

DEBONA, V. 2013. *A outra face do pessimismo*: entre radicalidade ascética e sabedoria de vida. 2013. Tese (Doutorado em Filosofia) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.

DEBONA, V. 2020. *A outra face do pessimismo*: caráter, ação e sabedoria de vida em Schopenhauer. São Paulo: Loyola.

DEBONA, V. 2022. Schopenhauer's great and small ethics: on the mysteriousness, (im)mediacy, and (un)sociability of moral action. *Schopenhauer-Jahrbuch*, Würzburg, Bd. 103, p. 57-80.

DEMIROVIÇ, A. 1999. *Der non-konformistische Intellektuelle*: Die Entwicklung der Kritischen Theorie zur Frankfurter Schule. Frankfurt am Main: Suhrkamp.

DURANTE, F. 2018. A esquerda schopenhaueriana no Brasil. *Voluntas: Revista Internacional de Filosofia*, Santa Maria, v. 9, n. 1, p. 137-147. https://doi.org/10.5902/2179378633548

DURANTE, F. 2022. *Entre heresias e atualidades de Arthur Schopenhauer*. Campinas: Editora Phi.

FAZIO, D. M. 2023a. La doppia faccia del pessimismo, *Cuadernos de Pesimismo*, México, n. 2, p. 109-123.

FAZIO, D. M. 2023b. Il male fisico e il male metafisico. Max Horkheimer: un eretico della Scuola di Schopenhauer. *Archivio di Storia della Cultura*, Napoli, anno XXXVI, p. 109-132.

HABERMAS, J. 1986. Bemerkungen zur Entwicklungsgeschichte des Horkheimerschen Werkes. In: SCHMIDT, A. (Ed.); ALTWICKER, N. (Ed.). *Max Horkheimer Heute:* Werk und Wirkung. Frankfurt am Main: Fischer Verlag.

HABERMAS, J. 2007. *Observações sobre o desenvolvimento da obra de Max Horkheimer.* Trad. Maurício Chiarello. *Revista Educação e Filosofia*, Uberlândia, v. 21, n. 42, p. 273-293, jul/dez 2007.

HABERMAS, J. 2022. *Teoria da ação comunicativa*: para a crítica da razão funcionalista. v 2.

Trad. Luiz Repa. São Paulo: Editora Unesp.

HORKHEIMER, M. 1968. *Kritische Theorie: eine Dokumentation*. Frankfurt am Main: Fischer.

HORKHEIMER, M. 1980. Teoria tradicional e teoria crítica. In: BENJAMIN, W.; HORKHEIMER, M.; ADORNO, T.; HABERMAS, J. *Textos escolhidos*. Trad. J. L. Grünewald et al. São Paulo: Abril Cultural.

HORKHEIMER, M. 2008. Notizen 1949-1969. In: HORKHEIMER, M. *Gesammelte Schriften*. Bd. 6. 2. Auflage. Hrsg. von Alfred Schmidt u. Gunzelin S. Noerr. Frankfurt am Main: Fischer.

HORKHEIMER, M. 2015a. *Teoria crítica: uma documentação*. Trad. Hilde Cohn. Apresentação Olgaria Matos. São Paulo: Perspectiva; Edusp.

HORKHEIMER, M. 2015b. *Eclipse da razão*. Trad. Carlos Henrique Pissardo. São Paulo: Editora Unesp.

HORKHEIMER, M. 2018. A atualidade de Schopenhauer. Trad. Lucas Lazarini. *Voluntas: Revista Internacional de Filosofia*, Santa Maria, v. 9, n 2, p. 190-208. https://doi.org/10.5902/2179378636126

HORKHEIMER, M. 2022. Kritische Theorie gestern und heute. In: HORKHEIMER, M. *Gesammelte Schriften*. Bd. 8. 2. Auflage. Hrsg. von Gunzelin S. Noerr. Frankfurt am Main: Fischer.

HORKHEIMER, M. 2023. *Taccuini 1950-1969*. Trad. Leonardo Ceppa. Bologna: Marietti.

HORKHEIMER, M. 2024a. Schopenhauer und die Gesellschaft (p. 43-54): die Aktualität Schopenhauers (p. 122-142), Pessimismus heute (p. 224-232). In: HORKHEIMER, M. *Gesammelte Schriften*. Bd. 7. 3. Auflage. Hrsg. von Alfred Schmidt u. Gunzelin S. Noerr. Frankfurt am Main: Fischer.

HORKHEIMER, M. 2024b. Die Sehnsucht nach dem ganz Anderen (p. 385-404), Zur Zukunft der Kritische Theorie (p. 419-434), Das Schlimme erwarten und doch das Gute versuchen (p. 442-479). In: HORKHEIMER, M. *Gesammelte Schriften*. Bd. 7. 3. Auflage. Hrsg. von Alfred Schmidt u. Gunzelin S. Noerr. Frankfurt am Main: Fischer.

HORKHEIMER, M. 2024c. Archiv der Goethe-Universität Frankfurt a.M. (UBA Ffm, Na 1, 762). Acesso em Julho de 2024.

JAY, M. 2008. *A imaginação dialética: história da Escola de Frankfurt e do Instituto de Pesquisas Sociais, 1923-1950*. Trad. Vera Ribeiro. Rio de Janeiro: Contraponto.

LUKÁCS, G. 2020. *A destruição da razão*. Trad. B. H. Hess, R. Patriota, R. V. Fortes. São Paulo: Instituto Lukács.

LÜTKEHAUS, L. 1980. *Metaphysischer Pessimismus und „soziale Frage"*. Bonn: Bouvier V. H. Grundmann.

LÜTKEHAUS, L. 1985. Einleitung II: Pessimismus und Praxis. Umrisse einer kritischen Philosophie des Elends. In: EBELING, H.; LÜTKEHAUS, L. (Hrsg.). *Schopenhauer und Marx. Philosophie des Elends – Elend der Philosophie*. Frankfurt a.M.: Syndikat.

LÜTKEHAUS, L. 2006. Ist der Pessimismus ein Quietismus?: Überlegungen zu einer Praxisphilosophie des Als-ob. In: HÜHN, L. *Die Ethik Arthur Schopenhauers im Ausgang vom Deut-*

schen Idealismus (Fichte/Schelling). Würzburg.

LÜTKEHAUS, L. 2007. Esiste una sinistra schopenhaueriana? Ovvero: il pessimismo è un quietismo? In: CIRACÌ, F.; FAZIO, D. M.; PEDROCCHI, F. (a cura di). *Arthur Schopenhauer e la sua scuola*. Lecce: Pensa Multimedia.

MATOS, O. 1993. *A Escola de Frankfurt: luzes e sombras do Ilumismo*. São Paulo: Moderna.

MATOS, O. 1989. *Os arcanos do inteiramente outro: a Escola de Frankfurt, a melancolia e a revolução*. São Paulo: Brasiliense.

MATOS, O. 1990. Apresentação. In: HORKHEIMER, M. *Teoria Crítica: uma documentação*. Trad. Hilde Cohn. São Paulo: Perspectiva.

MELO, R. 2011. Teoria Crítica e os sentidos da emancipação. *Caderno CRH*, Salvador, v. 24, n. 62, p. 249-262.

MIGGIANO, P. 2017. Influenze schopenhaueriane nella "Sehnsucht" del giovane Horkheimer. *Voluntas: Revista Internacional de Filosofia*, Santa Maria, v. 8, n. 1, p. 84-115. https://doi.org/10.5902/2179378633732

POST, W. 1971. *Kritische Theorie und metaphysischer Pessimismus:* Zum Spätwerk Max Horkheimers. München: Kösel-Verlag.

PLÜMACHER, O. 1884. *Der Pessimismus in Vergangenheit und Gegenwart*: Geschichtliches und Kritisches. Heidelberg: Georg Weiss.

RAMOS, F. C. 2017. "Schopenhauer como otimista": considerações sobre um apontamento de Horkheimer. In: CORREIA, A; DEBONA, V.; TASSINARI, R. (Orgs.). *Hegel e Schopenhauer*. São Paulo: ANPOF, 2017.

RUGGIERI, D. 2015. Schopenhauer's legacy and Critical Theory. Reflections on Max Horkheimer's unpublished archive material. *Schopenhauer-Jahrbuch*, Würzburg, Bd. 96, p. 93-108.

SEMBLER, C. 2013. Teoría Crítica y sufrimiento social en Max Horkheimer. *Constelacines: Revista de Teoría Crítica*, Madrid, n. 5, p. 260-279.

SCHOPENHAUER, A. 1971. *Gespräche*. Hrsg. von Arthur Hübscher. Stuttgart-Bad Cannstatt: Fromman.

SCHOPENHAUER, A. 2008. *Sämtliche Werke*. Hrsg. von Paul Deussen. 16 Bdn. München: Piper Verlag, 1911-1941. In: SCHOPENHAUER, A. "Schopenhauer im Kontext III" - Werke, Vorlesungen, Nachlass und Briefwechsel auf CD-ROM (Release 1/2008).

SCHOPENHAUER, A. 2005. *O mundo como vontade e como representação*. Tomo I. Trad. Jair Barboza. São Paulo: Editora Unesp.

SCHOPENHAUER, A. 2012. *Sobre a ética*. Trad. Flamarion Caldeira Ramos. São Paulo: Hedra.

SCHOPENHAUER, A. 2015. *O mundo como vontade e como representação*. Tomo II. Trad. Jair Barboza. São Paulo: Editora Unesp.

SCHMIDT, A. 1974. *Zur Idee der Kritischen Theorie*: Elemente der Philosophie Max Horkheimers, München: Carl Hanser.

SCHMIDT, A. 1977. Die geistige Physionomie Max Horkheimers. In: SCHMIDT, A. *Drei Studien über Materialismus*. München: Carl Hanser.

SCHMIDT, A. 2021. *Schopenhauer e o materialismo*. Trad. Maria Lúcia Cacciola. São Paulo: Clandestina.

TAUBERT, A. 1873. *Der Pessimismus und seine Gegner*. Berlin: Carl Duncker's.

TÜRCKE, C. 2019. Horkheimer e as tentações da teoria crítica. Trad. Rosalvo Schütz, *Problemata: R. Intern. Fil.*, João Pessoa, v. 10. n. 4, p. 166-182.

VEAUTHIER, F. W. 1988. Zur Transformation der Pessimismus-Motive im Denken Max Horkheimers. *Schopenhauer-Jahrbuch*, Frankfurt am Main, Bd. 73, p. 593-607.

WIGGERSHAUS, R. 2002. *A Escola de Frankfurt*: história, desenvolvimento teórico, significação política. Trad. Lilyane Deroche-Gurgel. Rio de Janeiro: Difel.

Michel Foucault, teórico crítico?: uma interpretação a partir de Judith Butler

Michel Foucault, critical theorist?: a Judith Butler's interpretation

Daniela Cunha Blanco[1]
Universidade Federal do Rio de Janeir
danielablanco27@gmail.com

Abstract: O artigo pretende responder às críticas de Habermas a Foucault, a partir da análise da crítica em Foucault em um diálogo com Butler. Habermas nos coloca as seguintes questões: 1) se a continuidade da tradição crítica kantiana em Foucault – dos textos *O que é a crítica?* e *O que são as luzes?* – é viável e como se deve compreendê-la, tendo em vista a crítica ao sujeito transcendental de *As palavras e as coisas;* 2) se o pensamento das relações de poder em Foucault constituiria uma *criptonormatividade* para a ação. Nossa hipótese é que a interpretação de Butler de Foucault responde tanto à possibilidade da continuidade da crítica kantiana, como à presença de elementos da *teoria crítica* da Escola de Frankfurt em Foucault. Será preciso seguir, em primeiro lugar a retomada da crítica kantiana feita pelo autor e, em segundo lugar, a normatividade que Butler identifica em Foucault.

Keywords: crítica; Judith Butler; Jürgen Habermas; Michel Foucault; normatividade; poder.

Resumo: This article aims to respond to Habermas' criticism of Foucault, based on an analysis of critique in Foucault in a dialogue with Butler. Habermas poses the following questions: 1) whether the continuity of the Kantian critical tradition in Foucault – from the texts *What is Critic*? and *What is Enlightenment*? – is viable and how it should be understood, considering the critique of the transcendental subject in *The Order of Things*; 2) whether Foucault's thinking about power relations would constitute a *cryptonormativity* for action. Our hypothesis is that Butler's interpretation of Foucault responds both to the possibility of the continuity of Kantian critique and to the presence of elements of the *critical theory* of the Frankfurt School in Foucault. It will be necessary to follow, firstly, the author's resumption of Kantian critique and, secondly, the normativity that Butler identifies in Foucault.

Palavras-chave: critique; Judith Butler; Jürgen Habermas; Michel Foucault; normativity; power.

[1] O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. (Bolsa de Pós-doutorado do Programa Institucional de Pós-Doutorado (PIPD) da CAPES.

Recebido em 30 de junho de 2025. Aceito em 10 de novembro de 2025.

O debate entre Habermas e Foucault: o problema da crítica totalizante

No texto *O que é a crítica? Um ensaio sobre a virtude de Foucault*, Judith Butler (2013) se dedica a interpretar o sentido da crítica em Foucault e a nela defender uma ideia controversa, qual seja, que existiria uma normatividade no pensamento crítico de Foucault. Com tal defesa, a autora responde ao debate que Habermas traça com Foucault. Em *O discurso filosófico da modernidade*, Habermas (2002) aponta que, se por um lado, Foucault teria feito uma crítica à razão universal que, na mesma linha do que Adorno e Horkheimer já haviam feito em sua *Dialética do esclarecimento*, encontrava sua razão de ser nas consequências políticas de uma razão totalizante, por outro lado, a pretensão totalizante da crítica da razão de Foucault lhe retiraria qualquer possibilidade de pensar um espaço de ação política. Em outras palavras, se a razão em sua totalidade deve ser recusada, como regular o campo de ação para medir seus objetivos e efeitos? Ou, ainda, recusando a razão em sua totalidade, qual seria o paradigma capaz de embasar as ações para a emancipação? Sabemos que a resposta encontrada pelo próprio Habermas trata de propor um deslocamento da razão subjetiva para a racionalidade comunicativa, cujo caráter pragmático a localiza na história e não mais no espaço metafísico. A universalidade da razão deixa de ser aquela cujas consequências fascistas foram amplamente analisadas por Adorno e Horkheimer, e passam a dizer respeito à esfera pública, na qual as regras formais de seu funcionamento garantiriam a participação universal de todos. É justamente esse novo paradigma da razão que faltaria a Foucault.

No texto *Com a flecha dirigida ao coração do presente: sobre a preleção de Foucault a respeito do texto de Kant 'O que é Esclarecimento?'*, Habermas (2015) volta a criticar Foucault a partir do que aponta como uma contradição em seu pensamento. Tendo agora, após a recente morte de Foucault, tido acesso a seu texto *O que são as luzes?* – no qual Foucault retoma o texto de Kant sobre o esclarecimento a partir de uma interpretação que o permite afirmar seu próprio pensamento como uma continuidade em relação à crítica kantiana –, Habermas questiona como é possível conciliar o diagnóstico de Foucault sobre o pensamento de Kant em *As palavras e as coisas* com essa nova interpretação de *O que são as luzes?*. Se, diz o autor, em *As palavras e as coisas*, Foucault havia diagnosticado em Kant a relação intrínseca entre a analítica da finitude e uma ideia de progressão infinita do conhecimento, afirmando tal relação como uma vontade de saber ligada a um poder, o que aparece em sua interpretação de Kant a partir de *O que é o esclarecimento?* seria bem diferente. Nas palavras de Habermas,

> Se antes Foucault farejara essa vontade de saber nas formações modernas do poder apenas para denunciá-la, ele a mostra agora sob uma luz inteiramente diferente: como o impulso crítico digno de conservação e carente de renovação, que vincula seu próprio pensamento aos começos da modernidade (HABERMAS, 2015, p. 198).

Em outras palavras, aquilo que teria fundamentado em *As palavras e as coisas* uma refutação radical de toda a racionalidade moderna, em *O que são as luzes?* teria se transformado em uma proposta de crítica da crítica que colocaria Foucault na esteira da modernidade. O que leva Habermas a questionar: "como a autocompreensão de Foucault como um pensador na tradição do Esclarecimento se concilia com a crítica inequívoca justamente a essa forma de saber da modernidade?" (HABERMAS, 2015, p. 195). Como, ainda, justificar que apenas a razão de Foucault, esta que apresenta uma leitura genealógica crítica das ciências humanas, permaneceria fora das relações de poder e de dominação da razão?

Habermas nos deixa as seguintes questões em aberto: em primeiro lugar, trata-se de compreender se essa continuidade da tradição crítica kantiana em Foucault é, de fato, viável, e como se

deve compreendê-la, sem que seja preciso abandonar a crítica que o autor havia feito ao sujeito transcendental em *As palavras e as coisas,* e, em segundo lugar, e mais importante, trata-se de questionar se a crítica de Foucault às relações entre poder e saber nos deixaram em uma completa ausência de normatividade para a ação. Em nossa interpretação, essas duas questões fundamentam a leitura que Butler faz de Foucault quando afirma não apenas a continuidade da crítica no autor, como, à revelia dele, a existência de uma normatividade em seu pensamento. Nossa hipótese é a de que a interpretação de Butler em torno do pensamento da crítica em Foucault tenta responder não apenas à possibilidade da continuidade da crítica kantiana em Foucault, mas, também, a continuidade da *teoria crítica* em Foucault, aqui já nos referindo ao modo com que ela aparece no século XX, seja com a primeira geração da Escola de Frankfurt, com Adorno e Horkheimer, seja com a segunda geração, da qual Habermas é um dos principais expoentes.

Será preciso analisar como Butler constrói essa difícil conexão entre a crítica enquanto crítica a qualquer pretensão universalizante da razão e do poder (tal qual aparece nos trabalhos arqueológico e genealógico de Foucault) e uma pretensão reguladora da ação política, que corre sempre o perigo de recolocar o universal lá mesmo onde ele foi criticado. Será central, antes de mais nada, compreender a concepção de crítica em Foucault, traçando as linhas de ligação e de ruptura que ela traça em relação à tradição crítica kantiana. Essa linha, como veremos, será mais fácil de traçar, tendo em vista que o próprio Foucault (2011, 2002, 2008, 2017) se posiciona em uma continuidade – mesmo que crítica – da crítica kantiana, o que aparece já no texto sobre a antropologia de Kant, e irá reaparecer tanto em seu trabalho arqueológico de *As palavras e as coisas*, como em suas fases genealógica e ética, em textos como *O que são as luzes?* e *O que é a crítica?*. O que significa dizer que a questão da crítica e o diálogo com Kant perpassa toda a obra de Foucault. Por outro lado, a afirmação feita por Butler de uma normatividade em Foucault nos coloca esse outro desafio, qual seja, questionar se existe um diálogo do pensamento de Foucault com a teoria crítica. E não se trata aqui de questionar se Foucault estaria lendo os teóricos críticos e neles se inspirando, mas, antes, nosso interesse é questionar se o pensamento crítico de Foucault não estaria assim tão distante da teoria crítica como Habermas havia apontado. Trata-se de questionar: seria Foucault uma voz importante do pensamento da teoria crítica, para além de sua relação com a tradição crítica kantiana?

O paradoxo da arqueologia e da genealogia: um rompimento e uma continuidade em relação ao esclarecimento?

Trataremos aqui do primeiro dos problemas apontados por Habermas no pensamento de Foucault, qual seja, aquele que questiona como teria sido possível que Foucault tenha, em sua arqueologia, identificado em Kant a expressão da razão totalizante que era preciso recusar, e, no texto posterior sobre o esclarecimento, ter se afirmado na continuidade da crítica desse mesmo esclarecimento de Kant. Em *O discurso filosófico da modernidade*, Habermas afirma que Foucault, com sua genealogia, quer colocar em andamento "um *discurso especial*, que pretende se suceder *fora* do horizonte da razão, sem ser, no entanto, absolutamente irracional" (HABERMAS, 2002, p. 429). Comecemos por lembrar que Foucault, em nenhum momento pretendeu se apresentar como a voz destoante, a consciência acima da realidade capaz de ver na razão aquilo que os outros não viam. Aliás, sua crítica à figura do intelectual mostra bem a pretensão de se manter distante dessa posição. O mesmo vale para o modo com que Foucault se afastou do pensamento da ideologia para aquele das práticas sociais. Nancy Fraser mostra esse desvio quando afirma que

> A genealogia foucaultiana do poder moderno estabelece que o poder afeta a vida das pessoas mais fundamentalmente por meio de suas práticas sociais do que por meio de suas crenças. Isso, por sua vez, é suficiente para descartar orientações políticas voltadas principalmente à desmistificação de sistemas de crenças ideologicamente distorcidos (FRASER, 1989, p. 18, tradução nossa).

Nesse sentido, fica claro que a crítica de Foucault ao que ele denomina de "tirania dos discursos englobadores, com sua hierarquia e com todos os privilégios das vanguardas teóricas" (FOUCAULT, 2005, p. 13), não pressupõe uma possibilidade de colocar-se fora da razão, como única consciência capaz de enxergar a realidade por trás do véu da ideologia. Antes, o gesto crítico de Foucault é aquele capaz de inserir-se nas lutas e embates que tem lugar na própria constituição do saber e da racionalidade. Seria, portanto, mais proveitoso analisar o modo específico com o qual Foucault interpreta essa racionalidade totalizante a partir das relações de poder. Se é possível realizar uma crítica ao poder é em seu interior, com suas próprias operações e, para isso, é preciso compreender as formas de operações do poder – trabalho, justamente, de sua crítica.

Trata-se de compreender aquilo que Foucault (2005) afirma: que não existe o poder, como uma instância fixa, e sim que o poder existe em ato, não como algo que se possui, mas algo que se exerce. Nesse sentido, toda a crítica de Foucault à razão não diz respeito a uma razão que existiria como instância fixa, da qual seria possível escolher fazer parte ou colocar-se de fora. O que significa dizer que não se trata de criar um discurso que se coloque fora da razão, mas sim um discurso que, em seus gestos e atos, desloque as relações usualmente estabelecidas entre razão e poder. Nas palavras do próprio Foucault,

> A genealogia seria, pois, relativamente ao projeto de uma inserção dos saberes na hierarquia do poder próprio da ciência, uma espécie de empreendimento para dessujeitar os saberes históricos e torná-los livres, isto é, capazes de oposição e de luta contra a coerção de um discurso teórico unitário, formal e científico (FOUCAULT, 2005, p. 15).

A genealogia é, portanto, um outro modo de exercer a razão cujo intuito passa longe da busca por uma unidade discursiva que, por um lado, se mostra capaz de remeter todo acontecimento a uma causa única e, por outro, nesse mesmo gesto, aprisiona todo acontecimento à posição de mero efeito no interior de uma racionalidade global. Que a genealogia torne livres os saberes históricos, significa que ela os retira dessa relação de causa e efeito que engloba tudo em uma totalidade; significa, ainda, que a genealogia possibilita olhar para esses saberes em sua positividade – um conceito central para o pensamento de Foucault, desde sua arqueologia, que pretende dar conta da ideia de um saber que é produtivo e não mera expressão da infraestrutura econômica, como a ideologia é pensada[2].

Nesse sentido, é possível afirmarmos que o trabalho construído por Foucault (2002), já em sua arqueologia – que apontava para Kant como expressão das pretensões totalizantes do discurso da modernidade – não pretendia colocar-se fora desse discurso, mas, antes, compreender suas operações de poder e nelas se inserir. Se mais tarde, em textos como *O que são as luzes?* e *O que é a crítica?*, Foucault (2008, 2017) irá retomar Kant para se afirmar na continuidade desse discurso filosófico é com o intuito de nele encontrar as brechas, os vazios e as funções que

[2] Roberto Machado nos fornece uma interpretação desse tema ao afirmar que "uma grande novidade dessa pesquisa foi não procurar as condições de possibilidade históricas das ciências do homem nas relações de produção, na infraestrutura material, situando-as como uma resultante superestrutural, um epifenômeno, um efeito ideológico. A questão não foi relacionar o saber – considerado como uma ideia, pensamento, fenômeno de consciência – diretamente com a economia, situando a consciência dos homens como reflexo e expressão das condições econômicas. O que faz a genealogia foi considerar o saber – como peça de um dispositivo político que, como tal, se articula com a estrutura econômica. Ou, mais especificamente, a questão da genealogia foi a de como se formaram domínios de saber a partir de práticas políticas disciplinares" (MACHADO, 2006, p. 176).

podem ser operadas no sentido da resistência política ao poder. Isso significa dizer que o pensamento de Foucault, que afirma que onde há poder, há resistência, não se encontra assim tão distante do pensamento de Habermas – ao menos não a partir daquilo que o próprio Habermas aponta como ausência de ação política em Foucault. Quando Foucault afirma que o poder existe apenas em ato, seu intuito é apontar para o fato de que o poder é uma relação de forças na qual todos podem intervir. Sua crítica totalizante à razão não implica em um abandono dessa razão, mas sim na análise das operações de poder que ela coloca em ação, para encontrar as possibilidades de resistência.

Habermas (2015, 2002), por sua vez, afirma que a crítica da razão de Foucault recalca um contradiscurso que é imanente à própria modernidade, e cujo trabalho de exposição é justamente aquilo a que o autor alemão se dedicou a fazer. Seria, porém, apressado afirmar a impossibilidade do que Habermas denomina de contradiscurso à modernidade em Foucault, já que este sempre pressupôs que a resistência também é imanente ao próprio funcionamento do poder. Mudam, é claro, os termos e a perspectiva a partir dos quais essas relações são pensadas. É importante, afinal, lembrar que, se para Foucault é Kant que fornece a expressão máxima da crítica a qual é preciso retornar, para Habermas, por outro lado, é Hegel quem fornece e inaugura o discurso filosófico da modernidade. E é ao funcionamento da dialética hegeliana que Habermas se refere quando afirma a existência de um contradiscurso imanente ao discurso da modernidade. Como crítico de uma ideia de razão enquanto consciência, interessado em deslocar a racionalidade para o campo comunicativo, Habermas se coloca na continuidade da crítica social dialética que vai de Marx à escola de Frankfurt. Se trazemos aqui essa linha de pensamento a qual Habermas se alinha para afirmar sua distância em relação a Foucault é porque ela fornece uma justificativa para a crítica que Habermas faz a Foucault, sobre a qual Nancy Fraser nos apresenta uma síntese:

> Essa tradição analisa a modernização como um processo histórico de dois lados e insiste que, embora a racionalidade iluminista tenha dissolvido formas pré-modernas de dominação e privação de liberdade, ela deu origem a formas novas e insidiosas próprias. O importante sobre essa tradição, do ponto de vista de Habermas, e o que a distingue da tradição rival na qual ele situa Foucault, é que ela não rejeita integralmente os ideais e aspirações modernos cuja atualização de dois lados ela critica. Em vez disso, busca preservar e estender tanto o "impulso emancipatório" por trás do Iluminismo quanto o sucesso geral do movimento em superar as formas pré-modernas de dominação — mesmo enquanto critica as características negativas das sociedades modernas (FRASER, 1989, p. 35-36, tradução nossa).

A partir da interpretação de Fraser, podemos perceber que o que está em jogo é que, para Habermas, o discurso filosófico da modernidade é a própria dialética, ou seja, o processo dialético entre o discurso e seu contradiscurso. Processo que, de fato, não existe em Foucault, cujo pensamento se distancia da dialética e passa a se referir à ideia de relações de poder. O gesto de Foucault de recusar a dialética, para em seu lugar colocar uma ideia de relações de poder, não resulta em uma recusa da existência de algo constitutivo desse regime de poder/verdade que possibilite formas de resistência. A resistência, afinal, não precisa ter a forma do contradiscurso pensada por Habermas. Nesse sentido, recusar a dialética, não significa recusar completamente a racionalidade moderna. Antes, significa recuperar dela o sentido da crítica, mas sem utilizar-se do mesmo enquadramento, do mesmo vocabulário, que pressupõe a existência de uma divisão entre legitimidade e ilegitimidade (do poder, do saber), cuja função é justamente aquela de assujeitar certas categorias de saberes. É esse vocabulário e esse enquadramento liberal do pensamento do poder que Foucault recusa, mas não a crítica em sua totalidade.

A questão central aqui, então, se torna perguntar se aceitamos ou não a interpretação de que

o discurso da modernidade é definido exclusivamente como um processo histórico de dois lados que Habermas afirma ser a dialética. Trata-se, ainda, de nos perguntarmos se aceitamos a dialética como fundamento do *Esclarecimento* e da crítica. Em outras palavras, o que está em jogo é uma disputa por aquilo mesmo que compreendemos como sendo o sentido da crítica. Será a dialética fundamental e indispensável na definição da crítica? Ou, antes, a dialética é uma resposta histórica específica para a definição da crítica, mas, que, de nenhuma maneira, esgota a possibilidade de outras definições? É sobre essas questões que Foucault (2017, 2008) se dedica a pensar em textos como *O que é a crítica?* e *O que são as luzes?*, e sobre os quais nos debruçaremos a partir do diálogo com o texto de Butler (2013), *O que é a crítica? Um ensaio sobre a virtude de Foucault*, também dedicado a pensar o que é a crítica. Veremos como Foucault retoma o pensamento de Kant para extrair dele um sentido da crítica que não aponta para a dialética como sua fundamentação, mas, antes, para a questão da emancipação ou de como não se deixar governar.

Podemos, assim, afirmar que não há sustentação para o argumento de Habermas contra Foucault, de uma suposta incoerência entre a crítica à razão totalizante de Kant em *As palavras e as coisas* e a afirmação da continuidade em relação à crítica kantiana dos textos da década de 1980. A crítica de Habermas não se sustenta, dado que Foucault não pressupõe um fora da razão, tampouco a impossibilidade de resistir ao poder. Se Habermas pressupõe um processo histórico de dois lados que o permite criticar a razão, salvando seu caráter emancipatório, o mesmo vale para Foucault, já que, ao substituir a dialética (de dois lados) pelas relações de poder, as possibilidades de resistência (emancipatórias) ao poder lhe são também imanentes. Veremos como é isso que aparece no modo com que Foucault interpreta a crítica transcendental kantiana a partir de um deslocamento para o campo histórico, que retira a fundamentação das condições de possibilidade do saber do campo da universalidade transcendente, para realocá-la na contingência e na imanência histórica.

Tendo como intuito responder à questão da normatividade colocada no embate entre Habermas e Butler, faremos um percurso constitutivo da crítica em Foucault a partir de duas linhas argumentativas diversas e complementares: em primeiro lugar, trata-se de buscar a continuidade da crítica kantiana em Foucault a partir da busca pelos limites do pensamento e, em segundo lugar, mostrar a continuidade da crítica a partir da problematização da atualidade. É só a partir dessa compreensão da crítica em Foucault que poderemos retornar ao argumento de Butler sobre a existência de uma normatividade no pensamento de Foucault que o conecta à teoria crítica. Sigamos, portanto, o primeiro fio, aquele dos limites do pensamento como fio que conecta o pensamento de Kant em torno da crítica ao pensamento de Foucault. Faremos isso tendo como principal ponto de apoio o texto de Foucault, *O que é a crítica?* – publicado em 1990, mas cuja conferência original fora pronunciada em 1978. A partir desse texto principal, retornaremos a Kant, para compreender aquilo que Foucault encontra no autor como base para seu próprio pensamento.

A crítica entre Kant e Foucault: a questão dos limites do pensamento

Em *O que é a crítica?*, Foucault (2017) começa por propor uma análise da história da crítica a partir de uma perspectiva específica: trata-se de compreender como ela se constitui em concomitância com uma ideia surgida do interior da Igreja Católica, qual seja, "que cada indivíduo, seja qual for sua idade, o seu estatuto, e isto durante toda a sua vida e até nos pormenores das

suas ações, devia ser governado e deixar-se governar" (FOUCAULT, 2017, p. 33). Está em jogo aquilo que Foucault (2007) já havia constatado no primeiro volume da *História da sexualidade*, mostrando como as formas de governo que nascem entre os séculos XV e XVI, inicialmente restritas à pastoral cristã, no século seguinte, se espraiam para todo o campo social, constituindo-se como uma nova forma de poder. Foucault opõe a ideia de um poder soberano, que se baseava no funcionamento da lei, a um novo poder cujo funcionamento se dá no campo das normas e da normatização. As caraterísticas desse poder são objeto da analítica do poder foucaultiana, que aponta para o modo com que se desdobra em duas perspectivas específicas: uma que se refere a um poder disciplinar – voltado para o controle individual dos corpos – e outra que diz sobre um biopoder – cujo foco principal são as populações. Se Foucault retoma essa ideia no texto sobre a crítica é para mostrar como, concomitante ao surgimento dessas técnicas de governo, forma-se uma questão essencial sobre como não se deixar governar. A partir daí constitui-se a ideia de uma atitude crítica, uma espécie de virtude ou ética, intrínseca ao próprio surgimento dessa nova forma de funcionamento do poder.

Tal perspectiva da construção da história da crítica aponta para uma questão central na definição da crítica em Foucault, qual seja, que não existe uma crítica universal, ou mesmo uma crítica em si. Se podemos falar de crítica é a partir da ideia de uma atitude crítica sempre conectada a uma técnica de governo. A crítica é sempre crítica de alguma coisa. Foucault afirma que da pergunta sobre "como governar?" (os Estados, as cidades, uma casa, o próprio corpo, o próprio espírito) – pergunta que fundamenta as novas técnicas de governo, que se espalham da pastoral cristã para todo o campo social –, nasce a pergunta sobre "como não ser governado?". É preciso, porém, compreender que essa pergunta não denota um desejo de não ser governado de todo, já que não existe uma crítica universal. Trata-se de uma atitude crítica incessante, ou como afirma Foucault, "uma questão perpétua que seria: 'como não ser governado assim, por esses, em nome desses princípios, em vista de tais objetivos e por tais processos, não assim, não por isso, não por eles?'" (FOUCAULT, 2017, p. 34). Ideia que forneceria a Foucault a primeira definição da crítica: "a arte de não ser governado de tal maneira" (FOUCAULT, 2017, p. 34).

Essa definição marca o primeiro ponto de contato entre Foucault em Kant, já que Foucault irá buscar uma aproximação com a definição que Kant (2011) dava a *Aufklärung*, no texto *O que é Esclarecimento?*. Kant afirmava o Esclarecimento como a saída do estado de menoridade, cuja definição "é a incapacidade de se servir de seu próprio entendimento sem a condução de outrem" (KANT, 2011, p. 23). A partir do retorno a tal definição, Foucault questiona como situar a crítica em Kant. Em outras palavras, se é o Esclarecimento que é pensado como uma atitude de não se deixar governar assim, o que significará a crítica em Kant? Foucault afirma que ela deve ser pensada a partir de uma outra questão, qual seja, "sabes até onde podes saber?" (FOUCAULT, 2017, p. 37). E é justamente essa questão que nos fornece uma primeira chave de leitura daquilo que conecta a crítica kantiana à crítica foucaultiana: trata-se da questão dos limites do pensamento.

Retornemos a Kant, como faz Foucault (2017), para lembrar como seu projeto crítico é composto por três partes diferentes, às quais correspondem três questões diferentes. A *Crítica da razão pura* tem como questão "O que eu posso saber?", que trata de encontrar os limites do conhecimento. A *Crítica da razão prática,* tem como questão "O que devo fazer?"; questão moral, que coloca os limites da ação e de seus efeitos. A *Crítica da faculdade do juízo* tem como questão

"O que me é permitido esperar?"; questão complexa que articula os conceitos da natureza da primeira crítica ao campo da liberdade e da moralidade da segunda crítica a partir da ideia de uma finalidade da natureza. Sem nenhuma pretensão de querer sistematizar aqui o pensamento de Kant, nosso intuito é perceber aquilo que Foucault encontra como fio comum às três críticas kantianas e que será central para seu próprio pensamento. Segundo Marco Antônio Sousa Alves, "Foucault retém aqui uma lição importante de Kant: a centralidade da reflexão sobre as condições de possibilidade do conhecimento e do pensamento. A crítica é, assim, nada mais que uma reflexão sobre os limites" (ALVES, 2016, p. 16). A questão fundamental que aparece nas três críticas é, portanto, aquela dos limites. As perguntas pelo que posso conhecer, o que devo fazer e o que me é permitido esperar tratam das condições de possibilidades do conhecimento e da ação, dos limites transcendentais do que posso conhecer, fazer e esperar. Deixemos, então, claro: quando nos perguntamos sobre os limites (da ação, do que posso conhecer e esperar) estamos nos perguntando sobre as condições de possibilidade, cuja constituição é transcendental, já que existe um *a priori* que condiciona a ação e o pensamento.

Foucault irá interpretar a questão dos limites no projeto crítico kantiano conectando-o ao Esclarecimento, compreendido como saída da menoridade. Lembremos a afirmação de Kant de que se nos encontramos em um estado de menoridade, é por nossa própria culpa, já que isso demonstra uma falta de coragem para utilizar-se de nosso próprio entendimento e dispensar a condução de um outro. Foucault irá, então, afirmar que isso que Kant identifica como uma *coragem de saber*, que é invocada pelo Esclarecimento, "consiste em reconhecer os limites do conhecimento" (FOUCAULT, 2017, p. 38). Pode-se extrair daí que "Kant atribui à crítica, no seu empreendimento de dessubmissão relativamente à ação do poder e da verdade, como tarefa primordial, como prolegómenos a todo o *Aufklärung* presente e futuro, conhecer o conhecimento" (FOUCAULT, 2017, p. 38), em outras palavras, conhecer os limites do conhecimento. Com isso, Foucault interpreta em Kant uma ideia de Esclarecimento, ou seja, dessa atitude de não se deixar governar, como a expressão da atitude crítica que persistirá em seu próprio trabalho.

A questão central para Foucault é que não podemos nos esquecer que o Esclarecimento, compreendido como essa atitude crítica de não se deixar governar, é concomitante – e, ainda mais, só encontra seu sentido em resposta – ao desenvolvimento das diversas técnicas de governo surgidas com a nova forma de poder. Não por acaso, Foucault (2017) retoma em *O que é a crítica?* as ideias que já desenvolvera em *As palavras e as coisas* e em *A história da sexualidade*, livros nos quais o autor mostra (FOUCAULT, 2002, 2007b) que aquilo que se coloca como ocasião histórica para a necessidade de uma coragem de conhecer é o surgimento de uma ciência positivista autocentrada e o surgimento de um Estado que via a si mesmo como a própria racionalidade da história. Ou seja, o desenvolvimento desse entrelaçamento entre poder e saber na constituição de novas técnicas de poder se mostra como ocasião para o surgimento da questão sobre como não se deixar governar. O resultado disso é que as relações entre crítica e Esclarecimento tomam, cada vez mais, a forma de uma desconfiança ou de um questionamento constante: "de que excessos de poder, de que governamentalização, tanto mais incontornável porquanto se justifica na razão, não será esta própria razão historicamente responsável?" (FOUCAULT, 2017, p. 39). Ou seja, trata-se de questionar e de desconfiar do quanto a razão pode ser responsável pelos excessos de poder – uma desconfiança que é compartilhada, inclusive, com autores da teoria crítica da Escola de Frankfurt, como Adorno e Horkheimer. É, portanto, essa atitude crítica paradoxal, que é parte do Esclarecimento e dele desconfia, que Foucault trará para

seu próprio pensamento sobre a crítica, que o fará afirmar-se na continuidade do projeto crítico kantiano. Mas há, é claro, um desvio que será aqui operado por Foucault.

A *Crítica da razão pura*, texto que fornece as bases do projeto crítico kantiano, tem como intuito encontrar uma definição, uma determinação daquilo que podemos conhecer. Kant (2010) está interessado em encontrar os limites daquilo que nos é dado a conhecer, estando inserido em um contexto filosófico específico a partir do qual tenta solucionar o embate entre o empirismo e o racionalismo. A questão central é recusar que a origem do conhecimento seja inteiramente remetida às ideias imutáveis e universais, tanto quanto inteiramente remetidas à experiência empírica particular, propondo uma forma de reunir o particular e o universal. Para isso, o autor irá deslocar a universalidade do conhecimento do campo das ideias inatas para aquele das regras do conhecer, ou seja, do próprio funcionamento das faculdades cognitivas. Então, o que é universal e necessário em Kant não são as ideias, mas sim, as regras do conhecer, que podem ser aplicadas a qualquer experiência particular. O *a priori* em Kant, sendo justamente aquilo que fornece os limites do que nos é dado a conhecer, é, portanto, transcendental e necessário, fornecendo as condições de possibilidade do conhecimento a partir da universalidade do funcionamento das regras do conhecer.

Para compreendermos o desvio operado por Foucault em relação a essa constituição dos limites do conhecer em Kant como fundamento da crítica, será preciso retornar a seu trabalho arqueológico, em especial aquele de *A arqueologia do saber*, livro no qual o autor (FOUCAULT, 2007a) se empenha por explicar o método ou gesto metodológico que empregara em seus primeiros livros. O método arqueológico é aquele interessado em compreender o surgimento de um modo de pensar, de uma prática discursiva em uma dada época. Em outras palavras, o método arqueológico em Foucault se pergunta como foi possível, ou quais foram as condições de possibilidade do surgimento de um dado modo de pensar, ou do que o autor (FOUCAULT, 2002) denominara já no livro *As palavras e as coisas*, de episteme, conceito que aponta para um sistema de pensamento que conecta discursos e práticas, dando a ver um modo de pensar e perceber o mundo, predominante em uma dada época. Essa pergunta sobre as condições de possibilidade é, justamente, a pergunta sobre o *a priori*, ou sobre os limites para o surgimento de uma determinada episteme. A questão central aqui é que a pergunta sobre os limites em Foucault desloca o *a priori* de Kant, que era necessário e universal, e o afirma aqui como um *a priori histórico*. Então, as condições de possibilidade daquilo que conhecemos, em Foucault, não serão mais universais e nem necessárias, tampouco dirão respeito ao funcionamento das faculdades ou das regras do conhecer. O *a priori*, em Foucault, será histórico e, portanto, contingencial. Foucault não está interessado em pensar a continuidade eterna das regras do conhecer, mas, antes, a transformação histórica daquilo que nos é dado a pensar em uma determinada época. Marcando, portanto, não a continuidade de um funcionamento do pensamento, mas, antes, as descontinuidades históricas dos modos de pensar ou do que Foucault denomina de epistemes.

Os limites que Foucault encontra, portanto, não dizem mais respeito àquilo que podemos conhecer a partir de um funcionamento imutável das faculdades, mas, antes, àquilo que uma dada época nos diz ser possível conhecer. O que significa dizer que os limites apontam não para o que não podemos ultrapassar, mas, sim, para aquilo que é preciso ultrapassar a partir de uma atitude crítica. Se Kant buscava os limites para saber até onde podíamos ir, Foucault busca os limites para mostrar onde é possível lutar, recusando um saber que se mostra como universal e

imutável e mostrando-o em sua contingencialidade histórica, ou seja, mostrando-o como uma construção. Assim, a atitude crítica que Foucault constrói a partir de Kant não é a de ser capaz de julgar a partir dos critérios de uma razão subjetiva. Se o juízo é aquilo que nos fornece uma certeza, uma ideia de verdade, o que Foucault busca é suspender essas certezas e verdades e ainda compreender como se formam essas certezas e verdades, dado que as regras do conhecer são contingenciais.

Butler (2013) nos traz uma interpretação sobre os sentidos e efeitos desse deslocamento histórico da questão dos limites em Foucault. A autora afirma que se trata, agora, de colocar a seguinte questão: "que relação entre conhecimento e poder faz que as nossas certezas epistemológicas acabem servindo de suporte a um modo de estruturar o mundo que oblitera a possibilidade de ordenações alternativas?" (BUTLER, 2013, p. 162). Butler nos lembra, portanto, que o que está em jogo no modo com que Foucault interpreta uma ideia de crítica a partir de Kant são as relações entre poder e saber. E essas relações são aquilo que fundamentam o questionamento dos limites históricos de um saber, justamente para liberar os discursos silenciados por um poder-saber hegemônico.

Se, em Kant, encontrar os limites era encontrar as condições de possibilidade universais do que podemos conhecer, em Foucault, encontrar os limites é reconhecer a contingencialidade de nosso campo epistêmico com o intuito de transformá-lo. Trata-se de pensar como determinados saberes possibilitam o exercício do poder; como determinada ordenação do conhecimento fornece as bases para o estabelecimento de relações hierárquicas e de dominação; como o conhecimento dá as condições para que o poder se exerça de determinadas maneiras. É isso que é colocado em jogo por Foucault (2017, 2002, 2000, 1997, 2007b, 2010b, 2007c), não apenas no texto *O que é a crítica?*, como em todo seu projeto arqueológico (*As palavras e as coisas, História da loucura*) e mesmo na continuidade de seu trabalho genealógico (*Vigiar e punir, História da sexualidade*). O governo ou a governamentalização da vida, lembremos, se dá nesse nexo entre poder e saber, nessas técnicas de poder que se fundamentam em um saber. E é claro que essas formas de governo nos aparecem como as únicas possíveis, se mostram como necessárias, construindo a ideia de que existe uma única ordenação possível do mundo, uma única organização possível de nossas vidas. Se, como a crítica pensada por Foucault demanda, queremos não ser governados assim, desse modo, por esses princípios, precisamos desmontá-los, ou ao menos, desmontar seu caráter de necessidade e de universalidade. O que torna isso possível é, justamente, essa atitude crítica de exposição dos limites do campo epistemológico, desde que se compreenda que expor os limites do conhecimento é mostrar que ele é histórico e contingencial, e não necessário ou universal; é, portanto, desafiar o campo epistemológico, abrindo espaço para o pensamento de uma ideia de resistência. Nas palavras de Alves, "a crítica que antes identificava as limitações necessárias da razão humana, agora, com Foucault, assume a forma de uma possível transgressão, de outra maneira de pensar que pode emergir" (ALVES, 2016, p. 16). A crítica, que em Kant, era uma saída do estado de menoridade, será entendida por Foucault como uma atitude incessante de não se deixar governar. E as operações a partir das quais essa atitude é colocada em movimento é aquela da análise do encontro entre poder e saber com o intuito de desnaturalizar as relações aí estabelecidas. Com isso, podemos afirmar que encontrar os limites passa a significar encontrar os pontos de embate com a contingencialidade de um saber que se mostra como necessário em seus efeitos de poder.

Refletindo sobre a questão dos limites no gesto crítico de Foucault, Butler afirma:

> Indagamos pelos limites dos modos de conhecimento porque nos deparamos com uma crise dentro do campo epistemológico em que vivemos. As categorias segundo as quais nossa vida social é ordenada, produzem uma certa incoerência ou domínios inteiros de ininteligibilidade. E é a partir de então, a partir do esgarçamento do tecido de nossa rede epistemológica, que a prática da crítica emerge juntamente com a consciência de que nenhum discurso aqui é adequado e de que um impasse foi produzido por nossos discursos dominantes (BUTLER, 2013, p. 164).

Isso significa dizer que percebemos uma incoerência entre nossos modos de vida e o regime de inteligibilidade no qual esses modos de vida são acolhidos ou rejeitados. Tudo se passa como se não tivéssemos linguagem ou pensamento capazes de dizer sobre nossos modos de vida. É isso que Butler (2013) denomina de uma crise dentro do campo epistemológico e que irá basear seu próprio pensamento em torno das subjetividades abjetas e das vidas não vivíveis, dando continuidade à atitude crítica de Foucault. A questão central aqui é que, ao afirmar que quando nos indagamos pelos limites é porque nos deparamos com uma crise, Butler está afirmando que não apenas a pergunta de Foucault sobre os limites diz sobre uma crise dentro do campo epistemológico em que vivemos, mas, que, a pergunta sobre os limites em Kant também já dizia respeito a uma crise em seu tempo. De algum modo, portanto, a pergunta sobre os limites diz respeito à atualidade de um tempo histórico, e essa é, justamente, a segunda chave de leitura a partir da qual compreendemos que Foucault retoma a crítica kantiana e sobre a qual nos debruçaremos agora.

A crítica como atitude de modernidade: a problematização da atualidade e a normatividade

Cerca de seis anos após a publicação do texto *O que é a crítica?*, Foucault (2008) retoma o tema, agora, no texto *O que são as luzes?*, no qual aprofunda suas ideias em torno da crítica e no qual é possível encontrar as bases para aquilo que Butler (2013) defende no texto *O que é a crítica? Um ensaio sobre a virtude de Foucault*: a ideia de que há uma normatividade no pensamento do autor. Pretendemos mostrar como a ideia de *problematização* da atualidade em Foucault pode ser uma via de resposta às críticas de Habermas, a partir da compreensão de que a questão da normatividade aponta, justamente, para uma relação entre crítica e atualidade. Em outras palavras, se a normatividade é um horizonte prático-político para a ação no presente, é a atitude crítica que a fundamenta. Será necessário, portanto, compreender como a suposição de uma criptonormatividade em Foucault aparece em Habermas, tendo como auxílio a interpretação de Nancy Fraser, em seus múltiplos diálogos com Habermas e Foucault.

A questão central da crítica de Habermas (2015, 2002), em relação ao que identifica como uma *criptonormatividade* em Foucault, aponta para a ausência de uma razão reguladora capaz de responder aos próprios julgamentos políticos feitos pelo autor. Fraser (1989), no livro *Unruly practices: power, discourse and gender in contemporary social theory*, se junta ao coro de Habermas na crítica a Foucault, nos fornecendo uma explicação mais clara sobre o problema da normatividade em questão. A autora mostra como Foucault teria se afastado de uma normatividade própria à modernidade, fundamentada no contrato social, aquela que era construída a partir de uma divisão clara entre poder legítimo – aquele cedido a um poder soberano – e um poder ilegítimo – que trata dos abusos de poder por aquele que deveria governar em nome do povo. Essa divisão apresentada pelo enquadramento do contrato forneceria uma razão reguladora que permitiria ver e julgar uma forma de poder como justa ou injusta, como benéfica ou má. Para

Fraser, quando Foucault abandona esse enquadramento liberal da análise do poder, não deixa claro qual seria, ou mesmo se haveria um outro parâmetro normativo a partir da qual é possível afirmar, por exemplo, que "'disciplina' é uma coisa ruim" (FRASER, 1989, p. 42, tradução nossa). Questão que se mostra importante, já que, para Fraser, é preciso saber responder se uma determinada teoria é capaz de fornecer aquilo que a autora considera as principais tarefas da crítica social:

> podemos questionar, por exemplo, se a retórica de Foucault realmente faz o trabalho de distinguir os melhores dos piores regimes de práticas sociais; se ela realmente faz o trabalho de identificar formas de dominação (ou se ela ignora algumas e/ou reconhece erroneamente outras); se realmente cumpre a função de distinguir entre formas frutíferas e infrutíferas, entre formas aceitáveis e inaceitáveis de resistência à dominação; e, finalmente, se ela realmente faz o trabalho de sugerir não apenas que a mudança é possível, mas também que tipo de mudança é desejável (FRASER, 1989, p. 43, tradução nossa).

Em suma, a normatividade demandada por Habermas e por Fraser questiona se há alguma forma de razão reguladora a partir da qual é possível fazer juízos políticos que embasarão ações políticas. Em outros termos, para ambos, é preciso que essa razão reguladora defina e determine, antes de mais nada, um sujeito da ação política que separe, claramente, suas práticas daquelas práticas julgadas como promotoras de relações de dominação. Para Habermas, bem como para Fraser, portanto, é preciso que a crítica social fundamente um desejo de transformação do presente que só pode se dar já tendo estabelecido de antemão uma separação entre as boas e as más práticas, entre as formas de dominação e as formas de resistência.

Se as questões colocadas para a teoria de Foucault por Fraser (que ressoam a crítica à criptnormatividade feita por Habermas) são justas, a maneira pela qual ambos os autores recusam a teoria de Foucault não nos parece assim tão justa. Se o que está em jogo em Foucault, como ambos os autores bem o perceberam, é uma reconfiguração da atitude crítica a partir da recusa ou transformação de seus princípios fundamentais, é de se surpreender com o fato de que ambos não tenham concebido que a própria ideia de normatividade aqui deveria ser repensada. E é justamente essa a proposta de Butler (2013) ao responder às críticas de Habermas contra Foucault com a afirmação da existência de uma normatividade no autor. Como a autora afirma:

> Meu propósito é marcar a distância entre uma noção de crítica que é caracterizada como normativamente empobrecida em algum sentido e outra – que espero oferecer aqui – que é não apenas mais complexa do que seus críticos, geralmente, supõem mas que tem, eu argumentaria, fortes compromissos normativos que se revelam em formas que seriam difíceis, senão impossíveis, de ler dentro dos atuais parâmetros de normatividade (BUTLER, 2013, p. 162).

Butler pretende, portanto, mostrar como para analisar o sentido da resistência política em Foucault é preciso não apenas compreender a forma complexa de atuação do poder no autor, como, também, compreender que o sentido possível de uma normatividade a partir daí não pode permanecer aquele pressuposto por Habermas e Fraser. Nosso intuito será, assim, mostrar como a normatividade que Butler pretende afirmar em Foucault se conecta com a maneira pela qual o autor estabelece um pensamento sobre a atualidade, em especial, a partir de suas noções de atitude de modernidade e de problematização.

Butler chama a atenção para uma afirmação de Foucault, no primeiro texto sobre a crítica, que poderia passar como algo menor. Lá, Foucault afirma que está refletindo sobre "a atitude crítica como virtude em geral" (FOUCAULT, 2017, p. 32). Butler parte dessa afirmação sobre a virtude – que, aliás, se mostra como um conceito com tamanha importância em sua interpretação de Foucault, que a virtude aparece no título do texto sobre a crítica em Foucault –, para refletir sobre o que pode significar essa crítica interessada em expor os limites do horizonte episte-

mológico. Se a racionalidade é aquilo que guia nossas práticas sociais mais cotidianas, ao expor os limites históricos dessa racionalidade, mostrando sua contingencialidade, Foucault estaria abrindo espaço para uma recusa a ser governado por ela. O que faz com que Butler conclua que há, no autor, uma oposição entre a obediência (deixar-se governar) e a virtude, sendo esta compreendida, justamente, como a atitude crítica de não se deixar governar (assim, desse modo, por essas regras, etc.). A questão central é compreender que a atitude crítica pensada por Foucault constitui uma relação problematizadora com o campo epistemológico no qual a racionalidade e as práticas sociais se formam. O que, em última instância, significa dizer que a crítica problematiza a própria atualidade. E é justamente essa ideia de uma problematização da atualidade que Foucault (2008), em *O que são as luzes?*, irá extrair do texto de Kant, *O que é esclarecimento?*.

Foucault retoma aqui a leitura do texto de Kant conectando a pergunta sobre o que é esclarecimento ao projeto crítico em geral de Kant, e o faz a partir de um recorte específico. Foucault mostra como o *Esclarecimento* é compreendido por Kant como o momento no qual a humanidade será capaz de fazer uso de sua própria razão, recusando submeter-se a qualquer autoridade. E a conexão que estabelece com a crítica se dá a partir da ideia de que ela define "as condições nas quais o uso da razão é legítimo para determinar o que se pode conhecer, o que é preciso fazer e o que é permitido esperar" (FOUCAULT, 2008, p. 340). Assim, definir as condições do conhecimento significa aqui definir o uso legítimo da razão e das faculdades do conhecimento. E se podemos falar em uma autonomia da razão em Kant, é a partir da definição dos princípios desse uso legítimo.

A questão é que o texto sobre o esclarecimento de Kant – compreendido em sua conexão com seu projeto crítico mais geral – aparece para Foucault como um gesto paradigmático da filosofia, que inaugura uma relação específica com a atualidade. O autor diz sobre o texto de Kant que essa é a primeira vez que um filósofo conecta dessa forma

> a significação de sua obra em relação ao conhecimento, uma reflexão sobre a história e uma análise particular do momento singular em que ele escreve e em função do qual ele escreve. A reflexão sobre 'a atualidade' como diferença na história e como motivo para uma tarefa filosófica particular me parece ser a novidade desse texto (FOUCAULT, 2008, p. 341).

E esse gesto kantiano será interpretado por Foucault como o que denomina de uma atitude de modernidade, ou seja,

> Um modo de relação que concerne à atualidade; uma escolha voluntária que é feita por alguns: enfim, uma maneira de pensar e de sentir, uma maneira também de agir e de se conduzir que, tudo ao mesmo tempo, marca uma pertinência e se apresenta como uma tarefa. Um pouco, sem dúvida, como aquilo que os gregos chamavam de êthos (FOUCAULT, 2008, p. 342).

E se essa é uma atitude que Foucault identifica em Kant, isso significa dizer que, ao reunir uma reflexão filosófica a uma reflexão histórica, voltada para a problematização da atualidade, Kant estaria fazendo já isso que o próprio Foucault faz com seu gesto arqueológico, ou seja, questionar os limites históricos do conhecimento. Portanto, aquilo que no primeiro texto de Foucault sobre a crítica identificávamos como um deslocamento feito pelo próprio Foucault, do funcionamento necessário das faculdades cognitivas para a contingência histórica do saber, aqui aparece já como algo que o própria Kant já operava. Foucault, ao identificar a atitude de modernidade em Kant, está dizendo que o próprio Kant já pressupunha um caráter histórico e contingencial da razão, que Kant já questionava a singularidade da atualidade, ou seja, sua constituição histórica única. Então, os limites que Kant buscava já eram, também, históricos, mesmo que tivessem um caráter universal. Se pensarmos que as condições da experiência em

Kant são o espaço e o tempo, compreendidos como ideias universais que dão as possibilidades do campo de experiência, mesmo que essas ideias sejam universais, o modo de experienciá-las terá um caráter contingencial e histórico. Essa dobra do pensamento de Kant operada por Foucault não é, é claro, gratuita. Era preciso, afinal, ao se afirmar na continuidade do projeto crítico kantiano, retirar-lhe todo caráter metafísico, deslocando-o para o campo histórico.

Com esse gesto, Foucault acrescenta o sentido da atitude de modernidade àquela definição da atitude crítica já estabelecida no texto *O que é a crítica?*. A crítica, assim acrescida desse outro sentido, será definida por Foucault (2008) a partir de três eixos centrais: 1) ela é uma atitude limite: a análise dos limites que, em Kant, mostrava o que não era possível transpor, aqui, mostra a ultrapassagem possível, dando a ver que todo conhecimento que nos é apresentado como necessário e universal é, na verdade, efeito de imposições arbitrárias; 2) ela é uma atitude experimental: ao abrir domínios de pesquisas históricas capazes de olhar para a atualidade não a partir de projetos globais que pretendem escapar dela, mas, antes, percebendo os pontos nos quais a transformação é possível, sem que ela tenha uma forma definida de antemão; 3) seu trabalho será sempre parcial e local: o que significa que será preciso abrir mão dos desejos totalizantes, tendo em mente que isso não significa abrir mão de uma sistematização para que a crítica funcione. Esses três eixos definem a atitude crítica compreendida como um gesto incessante, sempre a se recolocar, que constitui, em última instância, um método.

Foucault acrescenta, ainda, à atitude crítica, agora também atitude de modernidade, a ideia da problematização. Noção que não é sistematizada ou definida de maneira definitiva em seu pensamento, mas que aparece de maneira esparsa em alguns textos centrais de Foucault. Como é ela que nos fornece uma ideia da relação com a atualidade que o autor identificava já no gesto kantiano de indagação sobre o Esclarecimento, nos interessa compreender tal noção como base do que Butler identifica como uma normatividade em Foucault. O autor afirma que

> O que é preciso apreender é em que medida o que sabemos, as formas de poder que aí se exercem e a experiência que fazemos de nós mesmos constituem apenas figuras históricas determinadas por uma certa forma de problematização, que definiu objetos, regras de ação, modos de relação consigo mesmo. O estudo (dos modos) de problematizações (ou seja, do que não é constante antropológica nem variação cronológica) é, portanto, a maneira de analisar, em sua forma historicamente singular, as questões de alcance geral (FOUCAULT, 2008, p. 350-351, grifo nosso).

A problematização aparece aqui já como esse gesto capaz de indagar o presente, reunindo um questionamento histórico a um questionamento teórico. Ou seja, a problematização possibilita perceber a singularidade histórica da atualidade, o que significa dizer que é a partir das problematizações que percebemos como algo se torna um problema para o campo do pensamento. É esse o trabalho feito, por exemplo, na *História da sexualidade*, em que Foucault (2007b) analisará como as práticas e experiências em torno do sexo o constituem como um problema para o pensamento. Não por acaso, a ideia de problematização reaparece no segundo volume da *História da sexualidade*, livro no qual Foucault (2010b) afirma que a história do pensamento em torno do tema da sexualidade tem como tarefa "definir as condições nas quais o ser humano 'problematiza' o que ele é, e o mundo no qual ele vive" (FOUCAULT, 2010b, p. 17). Isso significa que, ao buscar construir a história da sexualidade, Foucault não estaria interessado em compreender as ideias ou ideologias de uma dada sociedade, que viriam mostrar uma interdição moral do sexo; antes, o que está em jogo é analisar "as *problematizações* através das quais o ser se dá como podendo e devendo ser pensado, e as *práticas* a partir das quais essas problematizações se formam" (FOUCAULT, 2010b, p. 18-19). Ao conectarmos essa ideia com as discussões sobre a

crítica empreendidas até aqui, depreendemos que se trata de buscar os limites, ou as condições de possibilidade históricas a partir das quais o sexo se torna objeto de pensamento. E não um objeto qualquer, mas um a partir do qual o ser se pensa a si mesmo. A problematização do sexo é, também, portanto, uma problematização do ser em sua atualidade.

O conceito de *problematização* reaparece, ainda, no texto de 1984, *Polêmica, política e problematizações*, no qual Foucault (2010b) nos oferece uma pista que torna possível conectar a noção de problematização com aquilo que Butler identifica no autor como uma normatividade. O autor afirma a problematização como a "elaboração de um domínio de fatos, práticas e pensamentos que me parecem colocar problemas para a prática política" (FOUCAULT, 2010c, p. 228). O que não significa, continua o autor, buscar na política qualquer princípio que explicaria as experiências, mas, antes, analisar os problemas que essas experiências colocam para a política. A chave aqui é compreender que não se trata de pensar que o campo político define de uma vez por todas suas formas de expressão em outros campos, e que olhando para as experiências, seriamos capazes de remontar a esses princípios primeiros. Foucault recusa essa relação causal e entende que o conjunto de práticas e experiências de uma atualidade histórica elaboram o próprio sentido da atualidade, percebendo, com isso, a singularidade histórica como aquilo que coloca um problema para a política. A problematização, portanto, é um gesto de questionamento do campo político.

Mas é preciso ir além na compreensão disso que Foucault está aqui denominando de campo político. O autor afirma que a política não se refere "a nenhum desses 'nós' cujos consensos, valores, tradição formam o enquadre de um pensamento e definem as condições nas quais é possível validá-lo" (FOUCAULT, 2010c, p. 228-229). E afirma, ainda, que talvez fosse preciso se questionar se é preciso, de fato, colocar-se dentro de um *nós* prévio à questão para defender certos princípios e valores. Mas, se sua noção de problematização mostra como as práticas e experiências é que fazem nos questionarmos sobre o ser, é que fazem, antes mesmo, que o ser seja objeto de uma elaboração, é preciso recusar a existência desse *nós* prévio à própria elaboração. Com isso, Foucault propõe que o *nós* seja sempre um resultado provisório da questão, ou seja, que o sujeito da política resulta da própria problematização da atualidade, que resulta das práticas e experiências e de como elas colocam um problema para a política. Nesse ponto, Foucault faz uma espécie de aceno silencioso para Habermas, ao afirmar que essa concepção de um *nós* como resultado da problematização teria causado um certo incômodo para os leitores – e veremos como um desses leitores subentendidos é Habermas. O incômodo resulta "de saber se era possível construir um 'nós' a partir do trabalho feito e que fosse capaz de formar uma comunidade de ação" (FOUCAULT, 2010c, p. 229). Essa é, justamente, uma das críticas de Habermas a Foucault, qual seja, que a inexistência de um sujeito anterior à ação, já que ele é construído socialmente, impediria pensar em qualquer forma de ação política. Quem age, afinal, se o sujeito resulta ou se constitui na própria ação? Ou, nos termos aqui discutidos, que podemos pressupor um espaço de resistência se inexiste um *nós* da ação política, se o sujeito político só pode ser pensado como resultado da própria ação?

Para responder a tal questão, recorreremos às interpretações feitas por Butler em torno do pensamento de Foucault. Não apenas porque temos como interesse chegar naquilo que nos propusemos anteriormente – ver como a autora defende uma ideia de normatividade em Foucault como resposta às críticas de Habermas –, mas, também porque a autora nos parece realizar um

esforço constante por esclarecer e mesmo desenvolver pontos que permanecem como complicações no pensamento de Foucault. É com esse intuito que, em *A vida psíquica do poder: teorias da sujeição*, Butler (2017) analisa a constituição paradoxal da sujeição pensada como forma de poder em Foucault. A autora mostra como se trata de pensar que somos dominados por um poder que nos é externo, mas, que "nossa própria formação como sujeito, de algum modo depende desse mesmo poder" (BUTLER, 2017, p. 9). Butler afirma que, se entendemos que o poder sobre o qual Foucault fala é constituinte do próprio sujeito, determinando a condição de sua própria existência, ele deixa de se reduzir àquilo ao qual nos opomos, para ser também aquilo de que dependemos para existir. Em suma,

> a sujeição consiste precisamente nessa dependência fundamental de um discurso que nunca escolhemos, mas que, paradoxalmente, inicia e sustenta nossa ação. 'Sujeição' significa tanto o processo de se tornar subordinado pelo poder quanto o processo de se tornar um sujeito (BUTLER, 2017, p. 10).

Propondo que é preciso ir além de Foucault, e reunir uma teoria do poder a uma teoria da psiquê, Butler mostra como esse sujeito que se encontra na dupla situação de subordinado ao poder e constituído pelo poder é já o sujeito que está implicado na cena da psicanálise. Essa ambivalência do sujeito, em que sua autonomia é condicionada pela subordinação pode ser compreendida a partir da ideia psicanalítica de que "o sujeito surge em conjunção com o inconsciente" (BUTLER, 2017, p. 15). O que significa dizer que o sujeito surge já apegado àquilo ou àqueles de quem ele depende de maneira fundamental. Butler nos lembra como a psicanálise já mostrava que "a formação da paixão primária na dependência torna a criança vulnerável à subordinação e à exploração" (BUTLER, 2017, p. 15). E é justamente essa situação de dependência primária explicada pela psicanálise que irá condicionar "a formação política e a regulação dos sujeitos e se torna o meio de sua sujeição" (BUTLER, 2017, p. 16). Assim, o problema psicanalítico se transforma em uma via de resposta política ao paradoxo da sujeição foucaultiana. Porque Butler percebe que afirmar que o sujeito se forma na subordinação torna necessário assumir que a subordinação dá as bases ou as condições de possibilidade da continuidade do sujeito. Com isso, "desejar as condições da própria subordinação é, portanto, necessário para persistir como si mesmo" (BUTLER, 2017, p. 18).

Essa conexão entre uma teoria do poder e uma teoria da psiquê é aquilo que fornece as bases para responder à ambivalência da relação entre um sujeito construído e a possibilidade de ação política, assim descrita pela autora: "como é possível que o sujeito, tido como condição e instrumento da ação, seja ao mesmo tempo o efeito da subordinação, entendido como privação da ação?" (BUTLER, 2017, p. 19). Butler mostra como aquilo que denominamos de sujeito não pode ser confundido com o indivíduo. Este se transforma em sujeito ao ocupar, pela linguagem, uma posição no campo da inteligibilidade, ou seja, o campo das condições de possibilidade de sua existência e ação. O sujeito é uma ocasião para que o indivíduo atinja suas condições de existência e ação. Assim, se podemos dizer que existe uma ação de um sujeito que não pré-existe a ação, é a partir desse mecanismo circular, que compreende que o sujeito se constitui no ato mesmo pelo qual as condições de possibilidade de sua existência se formam. Por isso, é possível dizer, como Foucault dizia, que se tratava de pensar um nós da política, ou uma comunidade de ação política, como um resultado precário e parcial e não como aquilo que precede tudo. Por isso, é possível pensar que a problematização é o gesto mesmo pelo qual o sujeito se constitui e o nós, compreendido como sujeito político da ação, se forma.

Essa discussão sobre a constituição de um *nós* que não pode preceder à ação é justamente aq-

uilo com o que Butler (2013) inicia o debate de Foucault com Habermas, afirmando a existência de uma normatividade em Foucault. Apesar da autora não se referir ao texto de Foucault aqui citado – *Polêmica, política e problematizações*, no qual esse nós como resultado da problematização aparece –, é de se supor que a autora trace aqui um diálogo com ele, tendo em vista tanto a sugestão silenciosa que Foucault faz aí àqueles (Habermas) que não entendem como um sujeito constituído pode ser também um sujeito da ação, quanto a maneira pela qual a solução da Butler sobre a normatividade em Foucault aponta para a discussão empreendida em torno do nós como efeito da problematização.

Butler (2013) afirma que a crítica de Habermas sobre a ausência de uma normatividade em Foucault teria se mostrado acrítica em relação ao próprio sentido da normatividade. Nas palavras da autora,

> a questão 'o que podemos fazer' pressupõe um 'nós' que já se formou e que se conhece, cuja ação é possível e cujo campo de atuação é delimitável. Entretanto, se essas próprias formações e delimitações têm suas consequências normativas, será, portanto, necessário procurar pelos valores que sustentam o palco no qual a ação se desenrola. Essa procura será uma dimensão importante de qualquer investigação crítica acerca de questões normativas (BUTLER, 2013, p. 161-162).

Com tal argumento, Butler nos remete ao próprio sentido dado por Foucault à problematização, ou seja, ao modo com que uma determinada prática ou experiência se transforma em objeto de pensamento, constituindo um campo político, formando um sujeito e, portanto, as condições de possibilidade da ação. O que Butler mostra é que considerar uma ideia de normatividade que não se questiona sobre as próprias condições de formação do sujeito da ação e do campo de ação, é ignorar as múltiplas formas pelas quais o poder atua constituindo um campo de inteligibilidade no qual os modos de vida se dão. O que significa que é preciso levar em consideração que o poder atua construindo subjetividades que não possuem um campo de ação possível, que não se conhecem ou não são reconhecidos como sujeitos de ação.

Se é possível pensar uma normatividade em Foucault ela deve recusar, justamente, essa constituição prévia do sujeito da ação e do campo de ação. O que significa que aquilo que irá regular a ação, ou aquilo que servirá de legitimidade para sua teoria do poder terá, necessariamente, um aspecto de indeterminação – e talvez seja esse o incômodo de Habermas, para quem é preciso determinar a forma universal da ação comunicativa como parâmetro para corrigir as formas sob as quais ela opera. Habermas sabe que aquilo que pauta as práticas comunicativas entre os sujeitos são as dissimetrias, hierarquias e desigualdades, mas, defende que, ao orientarmos nossas ações pata o entendimento mútuo, ou seja, para a comunicação, antecipamos as condições ideias do agir comunicativo. Tendo e vista que elas não são cumpridas, nós identificamos as distorções e os obstáculos que impedem, a cada vez, a realização de uma ação comunicativa. De alguma forma, isso significa dizer que o projeto crítico de Habermas, esse que pensa as condições de possibilidade da experiência para delimitar um campo de ação, pressupõe uma certa comparação constante entre as regras do agir comunicativo e suas distorções, uma medição constante entre as regras ideais e a realidade. Isso nos mostra como, nas regras ideais, os sujeitos da comunicação já estão definidos, o campo de ação já está delimitado, os efeitos esperados pelo agir comunicativo já estão determinados. Já existe um nós, bem como as operações possíveis de comunicação e os efeitos esperados.

A questão central é que, se isso, por um lado, pode funcionar muito bem para aqueles que já possuem seu espaço de reconhecimento enquanto sujeitos em uma sociedade, por outro, coloca

um problema para aquelas vidas não reconhecidas, que são consideradas abjetas, para usar um termo de Butler (2010). Essas pessoas não fazem parte desse *nós* previamente construído, esse *nós* ideal. Do mesmo modo, as ações que lhes são necessárias não estão previstas, pois essas pessoas podem se encontrar muitas vezes em situações de completa exceção. Nesse sentido, a teoria do agir comunicativo de Habermas não daria conta de pensar aquelas vidas com as quais tanto Foucault quanto Butler estão preocupados. Se Foucault (2010a) já falava da vida dos homens infames, dos encarcerados, dos loucos, Butler (2010) desloca o debate para falar dos gêneros abjetos, das pessoas racializadas, dos imigrantes, dos em situação de guerra. O interesse por essas vidas – que em Butler aparecem sob a denominação de vidas não vivíveis – mostra como, se a esfera pública de participação já está definida, mas algumas subjetividades estão fora dela, não existe possibilidade de ação legítima para os excluídos.

A normatividade que Butler (2013) defende existir em Foucault deve dar conta de pensar um horizonte prático-político no qual o *nós* da ação nunca pode já estar constituído de antemão. Antes, esse horizonte deve prever, justamente, a constituição desse *nós* a cada cena política, a cada gesto de problematização. A questão central a partir da qual podemos compreender a constituição dessa normatividade está na própria definição da crítica como uma atitude interessada em encontrar e analisar os limites históricos do discurso e do saber, o que Butler denominou de limites da inteligibilidade. Mostramos até aqui como a crítica às pretensões totalizantes da razão em Foucault ao invés de localizá-lo fora do discurso da modernidade, o insere na continuidade crítica da crítica. Isso significa que a atitude crítica de Foucault é uma espécie de colocar-se nos limites do discurso moderno, é colocar-se, portanto, em uma certa indeterminação. Um espaço entre é, afinal, um espaço de indeterminação – que é o próprio espaço da sujeição sobre a qual Butler discorre. Mas esse estar entre é também um horizonte teórico que fundamenta um horizonte prático-político.

Podemos, portanto, concordar com Butler que há sim uma normatividade no pensamento de Foucault, mas que ela é pensada a partir da indeterminação. Já que interessa não termos um nós definido de antemão, um campo de ação já delimitado, interessa que a teoria que embase nossas ações abra espaço para a indeterminação, que nos possibilite olhar, a cada vez, para o tempo presente. Butler identifica um regime de inteligibilidade que nos possibilita viver, que em Foucault já aparecia como a normatividade constitutiva da vida. E essa normatividade diz respeito ao próprio modo de funcionamento do poder, criando limites do saber que se apresentam como necessários e universais. Se a loucura é, como Foucault (2000) nos mostra em *História da loucura*, constituída por um campo discursivo que a mostra como o limite necessário a partir do qual se está fora da razão, toda subjetividade constituída como louca se cristaliza em uma identidade excluída. Se toda forma de prática sexual que não se encaixa na norma, como o autor (FOUCAULT, 2007b, 2007c, 2010b) mostra nos três volumes de *História da sexualidade*, é identificada como anormal, isso significa que as subjetividades sexualizadas também são constituídas nesse limite discursivo que se apresenta como necessário. Então, o trabalho da crítica, como Foucault mostra, deve ser aquele de desnaturalizar esse saber necessário, mostrando que os limites nos quais ele se apoia são históricos e não necessários e universais. E é justamente essa busca constante por encontrar a indeterminação dos limites históricos que Butler (2013) irá identificar como a normatividade de Foucault. Afinal, se o poder atua por formas de determinações excludentes de limites, um horizonte prático-político capaz de responder a isso só pode passar pela busca das formas de indeterminação que podem desnaturalizar os espaços delimitados.

A atitude crítica que Butler encontra em Foucault constitui uma normatividade que diz respeito ao que a autora identifica na relação que Foucault estabelece entre uma ética e uma estética. Se a atitude crítica é aquela que diz de uma ética de um sujeito constituído nesse paradoxo da sujeição, a estética dirá respeito a uma prática da virtude compreendida por Butler como uma forma de "resistência a essa coerção que configura a estilização do 'eu' perante os limites estabelecidos do que se pode ser" (BUTLER, 2013, p. 172). Há nesse gesto um risco a partir do qual a própria ordenação do ser é levada ao seu limite. Isso que Butler identifica como uma prática da virtude em Foucault pressupõe essa espécie de risco, na medida em que esse tipo de prática

> postula para si um valor que não sabe como fundamentar ou assegurar, mas que é postulado não obstante, evidenciando, portanto, que há uma inteligibilidade para além da inteligibilidade já estipulada pela dupla poder-conhecimento (BUTLER, 2013, p. 176).

Muito distante, portanto, da normatividade esperada por Habermas, essa encontrada por Butler em Foucault diz respeito ao risco e a indeterminação de colocar-se nessa outra inteligibilidade recusada pelo campo normativo.

O que poderia nos levar a questionar se é possível concordar com a leitura de Butler sobre a normatividade, já que nos parece difícil aceitar que ela tenha o duplo sentido de uma regulação legitimadora para ação e de uma atitude crítica de indeterminação. Em outras palavras, se trata de questionar se é possível aliar regulação à indeterminação. Mas, se como mostramos, a forma de poder a qual Foucault responde é aquela da normatização, da criação de saberes necessários, de limites universais, temos que concordar que as únicas formas de resistência a esse poder devem passar por ultrapassar os limites da norma, por desnaturalizar o necessário, por tornar locais e particulares os limites. Ou seja, se trata de respostas que apontam, justamente, para formas de indeterminação. Mais do que nos perguntar, portanto, sobre uma possível incongruência dessa normatividade pautada na indeterminação, o que está em jogo é o risco ético que esse horizonte prático-político nos coloca. Trata-se, como afirma Butler, de

> estar nos limites da condição de reconhecimento: essa situação pode ser, dependendo da circunstância, tanto terrível quanto emocionante. Existir nesse limite significa que a própria viabilidade da vida de uma pessoa está em questão, o que podemos chamar de condições ontológicas sociais da persistência dessa pessoa. Também significa que podemos estar no limiar de desenvolver os termos que nos permitem viver (BUTLER, 2023, p. 47).

Isso significa que a atitude crítica coloca como normatividade para si mesma o incessante questionamento sobre a constituição de um nós estabelecido, de um campo de ação determinado. Aquilo que Habermas vê, portanto, como defeito na atitude crítica de Foucault, aparece para Butler, como sua virtude.

Considerações finais

Butler nos mostra como não é possível sustentar as críticas feitas a Foucault em relação a uma suposta impossibilidade de ação política de um sujeito construído. Foucault, em sua analítica do poder, já pressupunha possibilidades de resistência, já afirmava um espaço de ação incessante justamente como única forma de responder a um poder complexo que não podia mais ser resumido às operações de repressão e proibição. Mas, se é possível encontrar os caminhos para pensar a resistência em Foucault, a maneira de mostrar a resistência política como caminho de resposta às críticas de Habermas é mais difícil, tendo em vista que o autor não respondia de maneira direta e explícita a tais críticas. O que não nos impede de buscar os traços silenciosos desse debate nos textos de Foucault, como aquele, por exemplo, de quando o autor fala daqueles

que tem dificuldade de compreender a ação de um sujeito construído, sem se referir a Habermas como origem da crítica. Os motivos pelos quais Foucault recusava responder diretamente às críticas de Habermas são dados pelo próprio autor, no texto de que tratamos aqui, *Polêmica, política e problematizações*, no qual o autor afirma que se recusa a entrar em polêmicas, vendo nelas uma "figura parasitária da discussão e obstáculo à busca da verdade" (FOUCAULT, 2010c, p. 226). Aqueles que discutem a partir da posição da polêmica, diz Foucault, constituem um nós que é anterior à discussão, recrutando partidários para os quais o outro é um inimigo que deve ser derrotado. Críticas como as de Habermas a Foucault se configuravam para o autor como esse campo da polêmica, uma divisão dada de largada entre aqueles ditos pós-estruturalista e aqueles que pertenciam à teoria crítica. E essa divisão entre um nós e os outros, talvez, seja menos resultado do modo com que Habermas individualmente abordava o debate com Foucault, do que resultado de um momento histórico no qual os debates políticos tomavam uma forma específica de disputa no campo teórico. Disputa que, tendo já sido ultrapassada, pode ser repensada em um diálogo, de fato, para além da polêmica, como a própria Butler o faz em seu pensamento, ao colocar-se nesse limite entre a teoria crítica e o pós-estruturalismo. Desafiar os limites do campo epistemológico nos tempos atuais, como a autora propõe a partir de Foucault, talvez seja realizar esse gesto de aproximar dois campos de pensamento que nos pareceriam distantes, talvez seja afirmar Foucault como um teórico crítico, como um outro teórico crítico.

É importante lembrar que a defesa da existência de uma normatividade em Foucault não interessava ao próprio Foucault – que, é claro, se preocupava com as formas de resistência ao poder, mas não nos termos da normatividade. A questão interessa a Butler, justamente na medida em que a autora, por um lado, vê a necessidade de desnaturalizar os fundamentos e conceitos centrais da modernidade (e isso ela encontra em Foucault), mas ao mesmo tempo, ela tem pretensões mais próximas da teoria crítica. Então, Butler extrai de Foucault a possibilidade de desconstrução das relações entre poder e saber, mas também, torce o pensamento do autor para encontrar aquilo que interessa a ela: essa inserção na teoria crítica, que afirma um compromisso com o campo democrático a partir de um horizonte teórico-político para a ação, a partir de uma normatividade. Para encontrar esse horizonte em Foucault, Butler opera um gesto foucaultiano: o mesmo gesto que Foucault teria colocado em jogo em relação a Kant. Vimos como Foucault (2017, 2008) retorna ao texto *O que é esclarecimento?* de Kant para nele encontrar e afirmar como um interesse do próprio Kant a construção de um *a priori histórico*, ou seja, do pensamento das condições de possibilidade da experiência que são dados historicamente e, portanto, contingencialmente. Se Foucault lê Kant com esse interesse de dobrar o autor para nele encontrar um modo de encontrar a história e a contingencialidade, contra a transcendência e a universalidade, vemos como Butler (2013) encontra em Foucault um modo de dobrá-lo para encontrar uma possibilidade de pensar um horizonte prático-político para a ação, mas cujo teor recuse uma norma reguladora universal e pré-estabelecida. Com esses gestos, Foucault, de sua parte, reconstrói uma ideia de crítica como atitude voltada para o tempo histórico, parcial e local, enquanto Butler, por sua vez, reconstrói a ação política de um sujeito construído. A autora coloca em jogo um gesto para transformar uma atualidade na qual é preciso reunir a desnaturalização do saber a um horizonte prático para a ação política, ou seja, uma atualidade na qual é preciso e possível reunir os interesses do pós-estruturalismo, sem perder de vista a concretude da ação política com a qual a teoria crítica tanto se preocupa.

Desse modo, recorrer a Butler e a maneira com que a autora insere Foucault no debate que ele

recusara responder diretamente, nos auxilia a construir um jogo que, muito além de uma mera defesa irrestrita do pensamento de Foucault, trata de pensar as possibilidades de resistência reais que seu pensamento previa a um poder que se mostra cada vez mais complexo. Nesse sentido, Butler (2013) não apenas nos auxilia na análise do pensamento do autor, mas em como pensar avanços em relação a seu pensamento, construindo, por exemplo, uma ideia de política de alianças que, apesar de se basear na teoria do poder foucaultiana, vai além dela, em especial na maneira de pensar as formas de resistência. Como afirmam André Duarte e Maria Rita de Assis César (2019), Foucault e Butler concebem uma ideia de crítica como

> trabalho refletido da liberdade sobre si mesmo e sobre os outros, visando diagnosticar os múltiplos efeitos das relações de poder sobre os sujeitos, então ela exigirá dos/as resistentes a capacidade de se descentrar, de se transformar, de não permanecer sendo sempre os/as mesmos/as (DUARTE, CÉSAR, 2019, p. 36).

Duarte e César nos mostram, ainda, como essa concepção da crítica em ambos os autores fundamenta, em Butler, um "modelo da coalizão aberta entre diferentes movimentos minoritários, de modo a favorecer a formação de alianças políticas não fundadas em rígidas concepções de identidade" (DUARTE, CÉSAR, 2019, p. 36). Podemos dizer que as voltas que Butler dá com o pensamento de Foucault abre espaço para pensarmos as lutas dos movimentos políticos atuais para além de qualquer separação identitária, propondo, ao contrário, formas de coalização contingenciais que não precisam apagar as diferenças para serem colocadas em operação. E isso se torna possível pela maneira com que Butler desloca o pensamento de Foucault para posicioná-lo em uma proximidade com a teoria crítica.

Esse Foucault que aparece aqui como um teórico crítico, talvez seja o Foucault de Butler, mais do que o próprio Foucault. Mas se é preciso colocar-se na atitude crítica ou atitude de modernidade proposta por Foucault (2008), isso significa trazer para o jogo uma das características centrais definidoras da atitude de modernidade que Foucault buscara em Baudelaire, qual seja, aquela da ironia. Trata-se de um gesto de identificação daquilo que há de heroico na atualidade, mas que não pode passar por uma veneração cega. Não podemos, diz Foucault, ter a atitude do *flaneur*, que passa por todas as modas e novidades, de maneira acrítica. A ironia diz respeito a um jogo entre a verdade do real e o exercício da liberdade. Com isso, aquilo que parece natural a uma época, deve ser desnaturalizado e aquilo que é belo para uma época, deve ser compreendido a partir da construção histórica da beleza. Constatar a verdade heroica do real, desse real singular do tempo presente, tem sempre como intuito desnaturalizá-lo, ou seja, um jogo com o exercício da liberdade, aquela que nos possibilita imaginar. Nas palavras de Foucault,

> Para a atitude de modernidade, o alto valor do presente é indissociável da obstinação de imaginar, imaginá-lo de modo diferente do que ele não é, e transformá-lo não o destruindo, mas captando-o no que ele é. A modernidade baudelairiana é um exercício em que a extrema atenção para com o real é confrontada com a prática de uma liberdade que, simultaneamente, respeita esse real e o viola (FOUCAULT, 2008, p. 344).[3]

Essa é a atitude de Foucault, essa atitude crítica que identifica os limites do pensamento que definem uma época, mas não para destruí-lo completamente (imaginando uma utopia), nem para venerá-lo. Não seria esse também o gesto de Butler ao perceber no tempo histórico do

[3] Que a figura de Baudelaire, neste artigo, apareça apenas no final de nossa argumentação, não significa menosprezar a importância que o escritor tem na constituição de uma ideia de modernidade em Foucault. As escolhas feitas coadunam com o intuito de traçar o diálogo de Foucault com Habermas, Kant e Butler, para o qual outros aspectos dos textos de Foucault forneceram maiores contribuições. A aparição de Baudelaire aqui, a partir da ironia, trata mais de explicitar o gesto foulcautiano em relação à leitura de um Kant histórico e o gesto de Butler em relação à leitura de um Foucault normativo do que de compor a figura da modernidade e da crítica que nos interessava aqui.

debate de Foucault e Habermas o que lhe era singular, transformando-o a partir de um uso da liberdade? Ver em Kant uma preocupação com as condições de possibilidades históricas do saber, não era, afinal, captar o que havia de heroico em Kant, transformando em contingencialidade o que aparecia como preocupação com o universal? Um gesto, é claro, que não é ficção ou fantasia, já que, como vimos, Foucault encontra, de fato, no texto de Kant, uma preocupação com a história. Ou seja, Foucault transforma Kant, sem destruí-lo, pois o capta naquilo que ele é. Do mesmo modo, afirmar em Foucault uma normatividade, mostrando-o como um teórico crítico, não seria isso, também, uma transformação que não o destrói, mas que o capta no que ele é? Concordemos ou não com a ironia operada por Butler em relação a Foucault, ou mesmo com aquela operada por este em relação a Kant, fato é que a crítica se atualiza e se mostra mais capaz de responder ao tempo presente. Fiquemos, então, com o Foucault, teórico crítico, lido por Butler.

Referências Bibliográficas

ALVES, M. A. S. 2016. O homem e a crítica em 'As palavras e as coisas: Kant, Nietzsche, Foucault e além'. *Sapere Aude*, Belo Horizonte, v. 7, n. 12, p. 7-21, jan./jun.

BUTLER, J. 2017. *A vida psíquica do poder:* teorias da sujeição. Tradução Rogério Bettoni. 1ª ed. Belo Horizonte: Autêntica editora.

BUTLER, J. 2023. *Corpos em aliança e a política das ruas:* notas para uma teoria performativa de assembleia. Tradução Fernanda Siqueira Miguens e revisão Carla Rodrigues. 5ª ed. Rio de Janeiro: Civilização Brasileira.

BUTLER, J. 2013. O que é a crítica? Um ensaio sobre a virtude de Foucault. (G. H. Dalaqua, Trad.) *Cadernos de ética e filosofia política*, São Paulo, v. 1, n. 22, p. 159-179.

BUTLER, J. 2010. *Problemas de gênero:* feminismo e subversão da identidade. Tradução Renato Aguiar. 3ª ed. Rio de Janeiro: Civilização Brasileira.

DUARTE, A.; CÉSAR, M. R. de A. 2019. Crítica e coalizão: repensar a resistência com Foucault e Butler. *Aurora*, Curitiba, v. 31, n. 52, p. 32-50, jan./abr.

FOUCAULT, M. 2007a. *A arqueologia do saber*. Tradução Luiz Felipe Baeta Neves. 7ª ed. Rio de Janeiro: Forense Universitária.

FOUCAULT, M. 2010a. A vida dos homens infames. In: MOTTA, M. B. *Ditos e escritos IV: estratégia, poder-saber*. Tradução Vera Lucia Avellar Ribeiro. 2ª ed. Rio de Janeiro: Forense Universitária.

FOUCAULT, M. 2002. *As palavras e as coisas:* uma arqueologia das ciências humanas. Tradução Salma Tannus Muchail. 8ª ed. São Paulo: Martins Fontes.

FOUCAULT, M. 2005. *Em defesa da sociedade:* curso no Collège de France (1975-76). Tradução Maria Ermantina Galvão. São Paulo: Martins Fontes.

FOUCAULT, M. 2011. *Gênese e estrutura da Antropologia de Kant*. Tradução Márcio Alves da Fonseca e Salma Tannus Muchail. São Paulo: Edições Loyola.

FOUCAULT, M. 2000. *História da loucura.* Tradução José Teixeira Coelho Netto. 6ª ed. São Paulo: Editora Perspectiva.

FOUCAULT, M. 2007b. *História da sexualidade I:* a vontade de saber. Tradução Maria Thereza da Costa Albuquerque e J. A. Guilhon Albuquerque. 18ª ed. Rio de Janeiro: Edições Graal.

FOUCAULT, M. 2010b. *História da sexualidade II:* o uso dos prazeres. Tradução Maria Thereza da Costa Albuquerque e revisão J. A. Guilhon Albuquerque. 13ª ed. Rio de Janeiro: Edições Graal.

FOUCAULT, M. 2007c. *História da sexualidade III:* o cuidado de si. Tradução Maria Thereza da Costa Albuquerque e revisão J. A. Guilhon Albuquerque. 9ª ed. Rio de Janeiro: Edições Graal.

FOUCAULT, M. 2017. *O que é a crítica?* seguido de A cultura de si. Tradução Pedro Elói Duarte. Lisboa: Edições Texto e Grafia.

FOUCAULT, M. 2008. O que são as luzes? In: MOTTA, M. B. *Ditos e escritos II: arqueologia das ciências e história dos sistemas de pensamento*. Tradução Elisa Monteiro. 2ª ed. Rio de Janeiro:

Forense Universitária.

FOUCAULT, M. 2010c. Polêmica, política e problematizações. In: MOTTA, M. B. *Ditos e escritos V: ética, sexualidade, política*. Tradução Elisa Monteiro e Inês Autran Dourado Barbosa. 2ª ed. Rio de Janeiro: Forense Universitária.

FOUCAULT, M. 1997. *Vigiar e punir:* nascimento da prisão. Tradução Raquel Ramalhete. 16ª ed. Petrópolis: Vozes.

FRASER, N. 1989. *Unruly practices:* power, discourse and gender in contemporary social theory. Minneapolis: University of Minnesota Press.

HABERMAS, J. 2015. Com a flecha dirigida ao coração do presente. Sobre a preleção de Foucault a respeito do texto de Kant 'O que é Esclarecimento?'. In: *A nova obscuridade: pequenos escritos políticos*. Tradução Luiz Repa. 1ª ed. São Paulo: Editora Unesp.

HABERMAS, J. 2002. *O discurso filosófico da modernidade:* doze lições. Tradução Luiz Sérgio Repa e Rodnei Nascimento. São Paulo: Martins Fontes.

KANT, I. 2010. *Crítica da razão pura*. Tradução Manuela Pinto dos Santos. 7ª ed. Lisboa: Fundação Calouste Gulbenkian.

KANT, I. 2011. *O que é Esclarecimento?* Tradução Paulo Cesar Gil Ferreira e revisão Marco Antonio Casanova. Rio de Janeiro: Via Verita.

MACHADO, R. 2006. *Foucault, a ciência e o saber*. 3ª ed. Rio de Janeiro: Jorge Zahar.

The impossibility of blaming token people fairly: the problem of demands

A impossibilidade de culpar pessoas particulares de forma justa: o problema das demandas

Eduardo Estevão Quirino[1]
Universität Vechta
eduardoquirinio@gmail.com

Abstract: This essay attempts to do two things. First, to problematize the relation between obligations and demands. Second, to show that the popular principle of Ought Implies Can and a plausible reading of what it is for blaming to be fair are incompatible with some cherished assumptions to the point of being impossible to blame concrete people, those with flesh and bones, fairly. The argument can be summarized as follows: For a person to fairly blame another subject, they need to be justified in believing both that a) the subject was obliged to act in accordance with the demand associated to blame; and b) the subject was capable of acting in this way. Unfortunately, there are reasons to think that b) is never justified, leading to the blaming itself to never be justified. I try to show that this argument is almost entirely independent of positions on free will, making it the only overall skeptical argument (that I know of) that delivers this conclusion about fair blame and, consequently, moral responsibility to an extent, without involving substantial debates on free will. The essay connects some previously unassociated literatures on Ought Implies Can, blame, the nature of normative demands, Objectivism/Subjectivism about moral obligation, and moral psychology. The conclusion of the piece is not to be an endorsement that there is no fair blame, rather, it claims that these arguments should be taken as a *reductio ad absurdum*.

Keywords: blame; free will; moral epistemology; moral responsibility; normative demands; ought implies can.

Resumo: Esse ensaio busca fazer duas coisas. Primeira, problematizar a relação entre obrigações e demandas. Segunda, mostrar que o famoso princípio que Deve Implica Pode e uma leitura plausível sobre o que culpar de forma justa requer, são conjuntamente incompatíveis com algumas suposições muito aceitas, incompatibilidade essa que chega a ser impossível culpar pessoas concretas, de carne e osso, de forma justa. O argumento pode ser resumido assim: para uma pessoa poder culpar outra de forma justa, ela precisa estar justificada em acreditar tanto em a) o sujeito do julgamento era obrigado a aceitar as demandas associadas com a culpa, e b) o sujeito era capaz de agir de acordo com a demanda. Infelizmente, há motivos para crer que b) nunca é satisfeito, assim tampouco o é a culpa em si. Eu tento mostrar que esse argumento é quase inteiramente independente de qualquer posição sobre o debate sobre livre-arbítrio. Sendo dessa forma (que eu saiba) o único argumento para um ceticismo geral sobre responsabilidade moral que não depende de debates sobre livre-arbítrio. O ensaio conecta áreas antes não relacionadas que discutem Deve Implica Pode, culpa, a natureza das demandas normativas, Objetivismo/Subjetivismo sobre obrigação moral e psicologia moral. A conclusão cética, no entanto, não deveria ser aceita como tal, e sim como uma redução ao absurdo.

Palavras-chave: culpa; livre-arbítrio; epistemologia moral; responsabilidade moral; demandas normativas; deve implica pode.

[1] I thank all the people who discussed some of the contents of the present article with me in the presentations I made on it both at Vechta and at Curitiba. Prof. Jean-Christophe Merle, Frank Rettweiller, Tania Eden, Tales Yamamoto deserve special mention.

Recebido em 31 de julho de 2025. Aceito em 10 de novembro de 2025.

Introduction

There have been many controversies about the relationship between most kinds of actions and their respective morality. However, one such kind of action has seen scanter philosophical attention, namely, the acts of normatively demanding that others obey, or satisfy, some duties, norms, or expectations. I wish to add something about the moral significance of demanding itself, but because that would be a whole book (at least!) I will narrow down to a single kind of demand i.e. "blaming", that is, the demand agents make upon others (or themselves) to recognize and act upon their blameworthiness for a given misdeed.

If a person blames another they are making both a claim of fact (that bears justification); and a normative demand (that requires legitimacy). The claim of fact is usually something like: "you are blameworthy for the misdeed". The normative demand is usually (similar to): "you should recognize your error and act upon it accordingly to redeem yourself"[2], or "you should be morally reprehended". This second, normative part explains why I consider blaming an act of demanding.

In this paper, I argue that when it comes to the morality of demands, some unexpected tensions arise from our intuitions about fair blaming and the widely accepted principle 'Ought Implies Can (OIC)'. The tension is quite strong, in fact, for given some plausible empirical and philosophical assumptions, I show how we can conclude (absurdly, no doubt) that we cannot blame existent, flesh-and-bones people fairly. This conclusion is rather drastic, and should better be seen as a *reductio ad absurdum* against some views—exactly which ones is not clear. I will present a few options of how we could avoid this drastic outcome, and while I leave it opened which is the best, it is important to highlight that all options seem to come at a price. In other words, this argument is fun for the whole community of metaethicists.

For the sake of clarity, in Section 1, I will build my argument over SHER's (2006) account of blame, however, I do not think that much hangs on this account's details, for what is problematic about fair blaming seems to be core intuitions any plausible account of blame would have to endorse, or so I argue. In Section 2, I briefly present the discussion over whether the nature of obligations is subjective or objective and whether the answers are compatible with OIC. I also argue for a particular interpretation of OIC based on capability rather than nomological possibilities. Section 3 takes a detour and enters into some empirical claims about the biology of moral human agents. These empirical claims are more illustrative than actually crucial for the argument that follows, and are very general. Much the same argument could be advanced with merely philosophical thought experiments, but I think the more down-to-earth construction adds plausibility and shows that I am not just running after philosophical drivels.

These three first sections are rather brief, being here only to set the stage. Section 4 starts with an account of normative demands and their relation to obligations. Furthermore, I propose a definitions of Fair Blaming. Additionally, it presents the new concepts I introduce: "token/type people" i.e. flesh-and-bones, spatiotemporal people versus hypothetical, generalized people. The crux of the difference lies in the epistemic availability of morally relevant properties in the scenarios under consideration. When we are talking about type people, we can stipulate *all the morally relevant properties*, thus philosophers frequently build very precise examples to capture exactly some moral properties ignoring others. For token people, this is not so. It seems that

[2] Different theorists will have dispute which particular demands are being made by blaming. The one I chose is my own and a similar one, by SHER's (2006), will be discussed in section I.

the morally relevant properties in situations involving token people are a contingent, partially empirical matter. This section also explains how OIC is connected to the argument. Section 5 is where it all come together in the two *Epistemic Arguments Against Token Blaming*. Section 5 is the core of this paper. Finally, section 6 briefly explores what can be done about those arguments, showing some ways out the skeptical conclusion and their costs.

These arguments against fair blaming are the only arguments I know of for generalized skepticism against normative responsibility (epistemic and moral included) that has no dialectically relevant connections to free will or determinism. Other arguments are either locally skeptical, or depend on certain views on free will/determinism. My argument does depend on a certain view of free will, but it is a negative dependence (it depends on X view not being true) and X in question is such a wild belief about free will that not even Cartesian Dualists would think plausible. Hence, if not for anything else, the arguments might be interesting for that reason.

1. Blame and Blame realism

We start our discussion with a brief sketch of SHER's (2006) account of Blame. His account is very influential and is still well-taken, which partially justifies the narrow scope of the discussion I wish to advance. A methodological problem for me arises from one of the reasons his account is so distinctive: Sher is a masterful structural writer, such that each point of his is carved by the previous and lends itself naturally to the next. Although a beautiful piece of philosophy, this makes my work as a reconstructor of his account somewhat lacking, unless I devote myself to a long piece, which I cannot. Hence, the schematic version of his views I wish to advance should be taken more like a way to focus this discussion, rather than a proper reconstruction of his arguments.

With these remarks behind us, we can start exposing Sher's account. His work is divided in two parts. The first, chapters 2-4, are dedicated to answering the question of what justifies blamers (people ascribing blame) to blame subjects based on their past wrongdoing. The second part, chapters 5-7, hopes to answer more directly the question of what blame is. I will divide my presentation accordingly dealing with those early chapters in 1.1 and the later ones in 1.2.

1.1. Justifying Blame

Sher's arguments in part one start with a discussion about what justifies the transfer of opprobrium from a bad action to a bad actor. In Sher's view, genuine (moral) blame is only directed towards agents; whereas the reasons from which we ascribe blame are based on our reprobation of bad actions. Because we usually judge others based on their *past* actions, it is not immediately obvious why we should take action's badness to impinge upon the *current*, existing agent. If blame is to be justified, we must account for this apparent asymmetry (SHER, 2006, p. 7). He then follows up suggesting that one of the best extent proposals to deal with this problem is the Humean thesis, which is roughly as follows: we can infer from bad actions to the defect in an agent's character; and, because of this, we can blame the agent bearing the defective character for such. In other words, bad actions justify the blame only inasmuch as they offer good reasons to suspect bad character, and bad character is the main problem all along.

Sher's second chapter engages with this proposal and argues forcefully to its dismissal. He offers three examples, each of which brings forth a problematic objection. I will focus on the one I find most compelling. The account above is committed to the source of blame to be the badness of

character evidenced by bad actions. However, assuming a thoroughly evil person who has an opportunity to enact their evilness— cruelly taunting a child — the Humean account would suggest that the taunting is a legitimate reason to think that the aggressor deserves blame. All good. However, Sher points out that this is a problem, because the taunting only produced evidence for the claim that the aggressor was evil. Given that the person was already evil before the taunting, even if he had not acted in that way, he would already be blameworthy – his character being prior to his actions. But, Sher argues – correctly to my view –, that this is not acceptable. We should only be allowed to blame people after, and *because of*, their bad actions. (SHER, 2006, p. 30-31).

The Humean position being defeated, Sher rushes to point out which parts of it made it so plausible. He highlights two broad claims he thinks we should bring to bear in a plausible answer to the initial question. Claim (1) is that a person's *character* is uniquely well-situated to link the responsibility for the bad action to the agent's desert for blame. Claim (2) is that someone's bad behavior need not exhaust all the things for which one can be blamed (SHER, 2006, p. 7). The base reason for the first claim is that character, seen as a complex mixture of beliefs, desires, emotions and dispositions, is both co-occurrent with the agent, and causally responsible for their actions. This combination is rare enough, but his account will also benefit from another trait of characters: they seem to be especially connected to our self-identity: the thing we refer to when we say first-person pronouns like "I", "me" and "myself"[3]. More on this shortly.

In what comes to the claim (2), Sher's most sophisticated, and difficult to defend, position in the book is to maintain that we might be blamed for things we have no control over. Not actions, of course. If we have no control over an action, we have no blame for it (with plausible qualifications). But character traits are things for which, Sher argues, we can be responsible for even if we never have power to choose them. The argument is long and controversial (abduced in SHER (2006, chapter 4)). In particular, it is based on a reinterpretation of what it means for blame to be someone's.

He starts arguing that our self-identity is conceptually connected to our character. After that, he contains that our relationship to our actions is not conceptual, but causal. Thinking diagrammatically: we have a conceptual arrow connecting the self to the character, and a causal arrow connecting the character to the action and the follow-up outcome of this action.

The very popular condition of control for responsibility ensues regarding action, but not regarding our own character traits, because the ways in which actions are ours; and the way in which our characters are ours are different. That is, we need control over whether or not an action (or omission) happens because the connection between us and the action is only causal. If the causal link is broken, the action is no longer ours: it no longer bears on us. However, the link between us and our character is much more intimate. Our lack of control over how our character ended up being as it is does not touch upon the fact that it is *still our character*. In a nutshell: If it

[3] It is important to differentiate this from logical notions of identity, about which one may find PARFIT's (1971), and WILLIAMS' (1970) papers very insightful and still influential.

can be traced to us, we can be blamed for it (SHER, 2006, p. 58)[4]. This is clearly very polemical, yet so, I will grant it for the sake of argument[5].

To sum up the answer about justification: we wanted an explanation of what, if anything, can tie up the agent and their action such that their bad actions can lead us to blame them. The answer is that, differently from Humeans, Sher's account goes from the badness of the action (thus requiring it to have occurred, blocking above's objection) to the cause of the bad action, namely the vastly complex array of (among other things) beliefs, desires and dispositions that constitute the character of the person, which by its turn constitute the person. Hence, the bad action allows for the blaming of a given subject, because it was that subject's character that caused it.

1.2. The Nature of Blame

This subsection is more directly relevant to our discussion. Sher's answer here depends on the truth of the previous claim, to the effect that it is that connection that a person trying to blame must establish: if I want to blame you, I must believe that you (your character) is connected causally with the action for which I blame you. Hence, the first part of his account of blame is that blame involves *a belief* that the person being blamed acted contrary to the (presumed adequate/legitimate) standards, or that they have a character inclined to do so (SHER, 2006, p. 95). To narrow our attention in this discussion, we will only talk about moral standards, consequently, blame includes the belief of moral failure. However, this is not sufficient for a number of reasons associated to our everyday experiences. People might believe that the other person acted contrary to standards, but not blame someone, instead, the person might praise them for it, because the blamer might be against some widely accepted moral standard. To complement the account, then, Sher suggests that blame is a pair of dispositions, not any single one. The pair is the belief above added the desire that the blamed person had not acted in that way (or had not had that character) (SHER, 2006, p.103).

After suggesting his favored pair of dispositions, Sher shows that they can account for some desiderata for accounts of blame. His desiderata are basically that which we associate with blame in everyday life. Given that this whole discussion emerges from STRAWSON's (1962/2008) *Freedom and Resentment* paper, it is unsurprising that "reactive attitudes" take a major part in establishing what matters about blame[6]. I think it is rather straightforward to see that beliefs and desires suffice for most reactive attitudes, but two common attitudes we have associated to blame are less obviously connected to that pair i.e. apologies and reprimands.

Hence, to conclude this brief account, I wish to explain how Sher deals with these seemingly problematic cases, and make some claims about the metaphysical and epistemological implications of his account. So, regarding the first point, Sher argues that moral desires of the form above are distinct from the average desire one has because, given that morality is a public affair, externalizing our desires that morality should have been uplifted we present ourselves as moral, while, at the

[4] In fact, this formulation is similar to Sher's truism: "that people can only be blamed for what reflects badly on them"(SHER, 2006, p.58).

[5] But see SMITH (2013) for a criticism of his whole account, and SCANLON (2008, 2013) for a different account (also criticized by Smith).

[6] COATES & TOGNAZZINI (2013) explain the relation with Strawson's work and explore it in a helpful way. HIERONYMI (2020) offers a robust introduction and (re) interpretation of that complex piece.

same time, we avoid taking part on (or seeming complacent with) the transgression. The belief part is necessary for those two attitudes because on one hand, it explains why we apologize or reprimand. On the other, the belief that the person has acted poorly is a necessary part of a genuine apology or reprimand. If I apologize without believing that I have acted badly, I am faking my apology. If I reprimand you without believing that what you did was really wrong, I am merely scolding you. Sher's discussion goes on a while after this, but for the sake of brevity, I think this explanation suffices to show the gist of the solution.

Now, let us see the metaphysical and epistemological implications of his account. To do that, I must bring his general slogan for what a bad action is supposed to mean. A bad action is an action that acts against or ignores good moral reasons. If this account holds, then the belief associated with blame is the belief that the agent under evaluation has acted in ways that are contrary to good moral reasons either for ignorance or for inobservance of such reasons. The belief, therefore, has two main components: "X acted in a bad way" and "X ignored (culpably), or went against good moral reasons". This means that reasons against blame could operate by undermining either of the conjuncts. The first strategy would be the "Alibi claim" and the second the "Excusing claim", undermining the first and second conjuncts, respectively. The Alibi claim, naturally, is the strategy of arguing that it wasn't this person who did it. The Excusing claim, on the other hand, accepts that it was the person in question but wishes to point out that the bad action does not suggest moral unconformity. Because of this, the principle that Ought Implies Can (OIC) is a kind of Excusing claim. If it is invoked, it works by claiming that the person who did the bad thing, could not have obeyed the moral rule in question. This will be important later, for the excusing happens objectively and thus we might not know that someone was excused.

Insofar as the epistemological implications are concerned, the issue is more connected to the belief component. This epistemology of blame has two connected parts. The first regards the truth of the belief that the person being blamed is truly causally connected to the bad action (in adequate ways). Additionally, one must believe *justifiedly* that there are no sufficient excusing factors that mitigate, or fully exempt, the actor from blame. Fairly straightforward.

Now, for the metaphysical implications. As for the desire component, Sher argues that the desire we have that the person who broke a moral commitment had not done so (or that their character should have been better inasmuch as this moral commitment is concerned) is *constitutive* of fully accepting a moral stance. He argues, compellingly, that one cannot fully accept a moral commitment, stance, law, or etc. without at the same time desiring others (barring excuses) to follow it.

Regardless, I think the desire is associated to a normative demand too, because the desire Sher evokes has this constitutive role in making people part of the moral world. Desires, by themselves, seem to be the wrong kind of entity to explain this constitutive role entirely for they seem to be overly action-commanding, whilst moral acceptance is passive in a way: most of the time, the person just move along with it as if the moral norm was a piece of furniture.

Whereas the desire *plus* the demand could accommodate better the passive parts of fully belonging to the moral realm in a relatively minor fix, I think, moreover, that the demands for others and ourselves to follow the norms is *constitutively prior* to the relevant desire. The desire is only evoked *after* the moral norm was violated, and therefore cannot explain why the moral norm

exists in the first place. Hence, I think that the correct pair of intentions associated to blame is a belief and a demand, not a desire.

There is another metaphysical component here, that is tied to the beliefs, namely, their truth-conditions. Sher claims that our beliefs about the person have to be *true* about them if we are to blame appropriately, which suggests that blame has a fact of the matter. A very complicated one at that, given that the truthmaker associated to blame attributions corresponds to the existence of complex mental states causally connected to the bad action. Complex it may be, but real nonetheless. This realist position says that when a person believes others to be blameworthy, the type of belief is like any ordinary propositional belief. There is a fact of the sort "the person is blameworthy" that, if actual, would make my believing as much true.

In other words, anyone who posits a relationship between real states and blame, or between real obligations and blame, will be ascribing to a view that is committed to there being a fact of the matter that makes the claim "S is to be blamed for φ in t" to be either true or false. Adding OIC, we have the following:

BO (Blame for failing Obligations) If S is to be blamed for φ in t, then they ought to not have φ-ed in t.

Which implies (conceptually) that:

CC (Core Condition): If S is to be blamed for φ in t, then S could had not- φ-ed in t.

Focusing on Sher's account helps us see that CC is very plausible, but a glance to the very proposition should suffice to make much the same point. CC is the first assumption of the argument for generalized skepticism that I will advance. The next two are defended in the following sections.

2. Capability and Objective Obligations

Most accounts of blame are realists in the sense above. Now, it would be beneficial to think about what that entails for the nature of obligations. The reason being that CC only works in this realist way because the bridging principle appealed to, BO, applied an objective account of blame and obligations, such that whatever contradicts an obligation is blameworthy. This is seen as controversial for many. As I hope to build my argument on premises many hold, I should clarify what is the difference between subjective and objective ought, as well as point out why the later is to be preferred.

According to VRANAS (2007) and GRAHAM (2010), subjective obligations are those that (somehow) depend upon one's epistemic position regarding them. Objective obligations are those that do not depend thusly. Vranas articulates his influential account of OIC in terms of objective ought. I am objectively freed from an obligation if I (really) cannot dispatch it.

2.1. Possibility and Capability in OIC

'Cannot' here means the lack of capability, not the absence of possibility[7]. Although capability implies possibility, the converse is not true. Frequently, it is possible for a subject to do something while, nonetheless, they are incapable of doing it. For instance, a heavy metal door is locked from the inside and someone asks their friend to come open it. The friend would definitely be facetious

[7] For people interested in free will, this distinction is also why CC is not immediately a statement of the famous Principle of Alternative Possibilities criticized by FRANKFURT (1969).

if their answer was: "well, I opened the door earlier today, so it is certainly possible to open it. There is no reason for me to open it for you, you surely *can*(!) do it!". Yes, we can (in the sense of possibility) open the door. However, we cannot (in the sense of capability) open it from the outside, say.

Capability is, therefore, different than possibility. So, why should one prefer capability to possibility when formulating OIC[8]? An answer might benefit from definitions of those terms, and I will offer them later. However, we can run with an intuitive grasp for now, taking capability to be "something agents can do", and possibility to be "a way reality could be". I think the answer could be made in four short arguments. Those arguments are mine: Vranas only stipulates that capability is the best way to go, and KING (2019, ch.1) only alludes to the difficulties of separating the relevant kinds of possibility. A 'capability' understanding of OIC will be very important to my arguments in Section 5.

First, OIC is usually defended with an appeal to the unfairness of requiring people to do what they truly cannot; consequently, a good version of OIC would have to be related to this requirement. Unfortunately, possibility, seen as above, would be so broad as to not sustain this intuition at all. For instance, assuming only nomological possibilities (possible in accordance with the laws of physics), every invention and piece of knowledge humanity now has is nomologically possible to be had. So, assuming that having some knowledge might be a moral duty— to know how much medicine to give to a child is a moral duty for a parent needing to treat their child, say— we can conclude that everyone of those who were doctors in the past and who did not know about the importance of sanitation are blameworthy, because they ought to know about sanitation and it was *possible* for them to know it (the laws of physics do not prevent it, given that the laws are the same for us now, and we do know that). Therefore, OIC-possibility fails fairness: capability is to be preferred.

Second, duties are attached to agents; however, possibilities are attached to worlds. It is unclear why a duty that commands a person to do something would be undermined by things utterly unrelated to the agent. That said, this is reasonable for metaphysical and logical possibilities, but for reasons quite apart from OIC. Duties cannot be metaphysically or logically impossible, but that is not because OIC is true; rather, it is because duties are connected to the space of possible actions: they separate the actions one ought to do from those they not-ought (sic)[9]. Obviously, metaphysically impossible things are not part of the set of possible actions, likewise for logically impossible actions. This is a trivial claim[10]; and the relevant possibility for OIC must be nomological or natural possibilities. Now, it is not clear why the duties one must have would be connected, in any immediate way, to the laws of nature[11]. Contrarily, it seems rather plausible that the duty, being already connected to a given agent, would also be connected to other properties of that agent. This favors OIC-capability' accounts.

[8] Some people do not prefer capability to possibility, see BASSFORD (2022). For additional discussion on the formulation of an OIC thesis see: KING (2019, ch.1).

[9] For reasons related to precision, I had to follow Logic's (rather than English's) grammar here.

[10] Although see KING (2019, p.73), who argues that there could exist logically impossible ought. My argument is a simplified version of the one derived from WEDGWOOD's (2016, 2018) papers on the semantics of 'ought').

[11] This connection is so problematic that GOLDWATER (2020) uses it in an argument against OIC.

Third, the epistemology of capabilities is much easier to apply than the epistemology of nomological possibility. Although, it is not necessarily easy, in fact, much of my argument below is based on there being practically impossible cases. My contention is relative: what is bad for 'capability' is worst for 'possibility'. Modal epistemology is a field in and of itself, a particularly difficult one and most accounts it offers are, at best, unclear about how to deal with modality in general (they do tend to work better for logical possibilities, though). I do not wish to dwell too much on this[12], the point, however, stands: modal epistemology is a hugely controversial area; making any applications of OIC to be dependent upon it undermines its application considerably.

Finally, duties can be stronger and weaker, and one of the issues that surround how strong the duty is concerns how easy the person can dispatch the obligation. For instance, a person passing by a child who is in risk of being attacked by a small dog is strongly obligated to help the child. A person passing by a child in risk of being attacked by a bear is still obliged to help, but less so. OIC-capability helps explain this intuition. OIC-possibility does not, because possibility is a categorical thing. Either something is possible or it is not. This is not to say, however, that obligations are gradual all the way down. There may very well be a point from which a thing is, or is not obligatory, with no middle ground. But the strength of duties is a fairly intuitive notion that OIC-capability, but not OIC-possibility can make sense of.

I conclude that OIC should, as VRANAS (2007) suggests, be taken to be defined as capability. So now, let's define capability a bit more precisely according to VRANAS (2007, p. 170): Capability is the temporally indexed conjunction of ability and opportunity. Where ability is the set of skills, knowledge and bodily dispositions available for enacting some action; and opportunity is the condition external to the agential control that allows them to apply their ability. For instance, I cannot fly an airplane because I lack the knowledge and the skills necessary for doing so. Had I learned them, I would be able to fly an airplane. Yet, if there is no airplane, or the airplane is broken, I would not be capable of flying it— I would have the ability, but not the opportunity. So OIC means 'ought' implies the ability and the opportunity conjoined at the same time.

It is rather plain that capability is an objective matter. I may be capable to do something and not know it, or believe I am capable and not be. The sense in which it is objective is rather tricky, for one may not do some things they think they are incapable of doing, even if they are capable, thus never doing them which prevent a simple dispositional view of capability. So, the best way to capture this objectivity is counterfactually. The person S is capable to φ if had they tried in the opportune contexts, they would succeed (or have a plausible likelihood to succeed). Being as it may, the 'Can' in OIC is an objective feature of the world. It is also a descriptive and empirical one. Then, what about the 'Ought' part of OIC?[13]

2.2. The Objectivity of Obligations

Now, we need to delve into GRAHAM'S (2010) paper to discuss why we should consider obligations to be objective. Unsurprisingly, the discussion is rather complex. Before I set up the formal definitions, we can take a step back and assess the debate. What is at stake is, in a sense,

[12] But see BERGLUND (2005) for an at-the-time exhaustive overview and a still worthwhile introduction.

[13] I am citing only VRANAS (2007) because my aim is not in defending OIC, per se. His contributions to this discussion, however, surpasses that paper and cover almost all the literature in one way or the other. They are all great papers and deserve the read. See: VRANAS (2018a,2018b, 2024). Vranas is contrasted by the book by KING (2019) who argues against OIC.

where we should put the "moral camera" so to speak[14]. Should we follow the agent's perspective, or should we take a broad take on the whole scene and consider what would be the really best option? So, to keep the analogy, imagine the following scenario: we are watching one of those movies that the punch of the picture is to present a complicated moral situation. The same story was shot by the director in two different cameras (as far as I know, this is standard in the industry). In the first camera, we follow with the main character. We get some views as if it was first-person. In this situation, we see a fallen, broken chair with mud on it; we see the victim fallen down near the chair; and we see the knife with blood by the victims' side. Later on, we meet a person who was a suspect and they had the mud in their feet. They also had motive, and there was a whole talk about the person liking knives for whatever reason. Ok, the spectator is set to blame this person for the murder.

The other camera, however, keeps the broad perspective over the whole house. In this camera, we can see another person hiding in the living-room, and when chance allows, escaping. At that time, we also get a close in their feet touching the source of mud: a puddle outside the house that anybody coming in or passing by would have to step on. Oh, wow, the villain was not the suspect from before! The question now is: who will the main character blame?

Well, the terrible screenplay above suggests forcefully that it will be the suspect, not the villain. The ethical question comes now: who *should* the main character blame? (I am a bad screenwriter, so, if the above case seems insufficient for blame, just imagine that whatever the evidence needed for blaming the suspect is filmed in the first camera; and that a full explanation of those pieces of evidence is shown in the second camera, as well as this additional, rather fatal evidence that the villain was in the crime scene.) The point is: in the first camera, the main character is fully justified in thinking that the suspect is to be blamed. In the second camera, we get the truth: that another person did it. Now, if we are objectivists, we would want to say that the main character should blame the actual villain. If we are subjectivists, then we would hold that the main character is correct in blaming the suspect, even if this would be the wrong (factually incorrect) choice.

With this intro, we can see the definitions Graham offers for subjectivism and objectivism. Starting with the former:

> A moral theory, T, is ability-constrained-evidence-subjective=def. according to T, a person has the moral obligations that she has at a time solely in virtue of both facts about her abilities and facts about her evidential situation at (or prior to) that time (GRAHAM, 2010, p. 90).

And the latter is simply: "A moral theory, T, is objective=def. it is not the case that T is [ability-constrained-][15] evidence-subjective" (GRAHAM, 2010, p.89).

The definitions in that paper are done twice. The first to capture what subjectivists are saying. The second was to amend that definition to deal with a problem connected to taking OIC and subjectivism together. The problem is easy to see: the first definition of subjectivism did not include the ability restraint. So, if one believes in OIC, and that what they should do is based on the epistemic states of an agent, then, given that we can be wrong about our own capabilities, it follows that one could believe that they should φ even if they can't. Which contradicts OIC. To mend this problem, the definition quoted above was made. Graham points out some reasons for

[14] GREENE (2013, p. 133-1355) also uses the metaphor of the moral camera, although in a different context and with a different emphasis.

[15] Author's addition between brackets.

concern regarding that position. The subjectivist wants to hold that only the agent's epistemic stance can account for what they should do, *except* for their own ability. This ability is an objective feature in a subjectivist account. The only reason for the concession is the (reasonable) desire to keep OIC and subjectivism together. This is the definition of an *ad hoc* theoretical construction. Graham then asks: why not more objective conditions? And if more, why not *only*? (this last part is somewhat implied by his considerations).

I agree with this point. Subjectivism and OIC do not go hand-in-hand. If the agent, in a given point, was capable of getting more information, and, in those conditions, they were obliged to do that, OIC seems to warrant that the person *should get more information*. So, the fact that the person is not better informed is not a good reason to not *require* better information. If OIC is held, it is done so superfluously. If this is the case, then subjectivism would have to add a further argument for why someone's not *actually* knowing would entail that they *couldn't* know. The logical inference is obviously not valid, so they need additional reasons. If, on the other hand, they assume that not knowing does not prevent one from being obliged to know, then subjectivism seems untenable; given that if people could know better, and they should know better, then the fact that they do not know better is irrelevant for their moral situation, contrary to subjectivism.

Of course, holding OIC is not supposed to be the *only* excusing factor in the subjectivist theory. But the addition of other excusing factors must, at once, explain why not being better informed is a sufficient reason to forgo obligations to be better informed; furthermore, they would also maintain OIC saying that not being able to do something is sufficient for not being obliged to do it, independently of information people may have about their capability. Combining both claims we would have the following: what I know *now* limits what I should know; and what I cannot know limits what I should know. These combined propositions are not equivalent— but are dangerously similar— to a wholesale rejection that we have moral obligations to know things we do not currently know, or believe. I have not enough time to press the issue to its limits, but I hope that these remarks make it clear that any subjectivist accounts will have to account for this, and it is a prima facie objection to the combination of subjectivism and OIC. At best OIC is superfluous, at worst it offers a principle that claims a condition for exculpation that competes with subjectivism.

With the prima facie case delineated above, I move to Graham's two arguments for objectivism. First, we have the argument from correction, attributed to W.D. Ross's book "the Right and the Good". The argument is rather simple. Back to the two cameras above, according to the first-person camera, the main character should blame the suspect. In the second camera, they should blame the villain. The difference is that in the first, the main character has false, but justified belief. If this belief was corrected: going in accordance with the second camera, is it really true to say that their moral duties was changed? Isn't it more plausible that the person was *corrected*—implying that they were *wrong before*—rather than *redirected*— implying that the it was correct— but had to change for some reason? The argument suggests that when people go from imperfect knowledge to perfect knowledge, what change is not their moral obligations, but the correctness of their beliefs: they were supposed to act according to the corrected beliefs all-along[16]. This suggests that the objectively correct thing to do also matters and objectivism is correct.

[16] This argument has a question-begging ring to it. But so is true for most intuition-based arguments.

The second, and more persuasive, argument is, in Graham's own words:

> The second argument for objectivism goes as follows: the question I want answered when I ask myself what my moral obligations are is the same as that which I want answered when, in seeking your help, I ask you what they are; but, to adequately answer me you don't need to consider my evidence concerning my situation; therefore, my moral obligations don't depend on my evidence concerning my situation; so, objectivism is true (GRAHAM, 2010, p. 91).

Graham is very direct and succinct, warranting the direct quote. However, if I could suggest one amend: the adviser does not need *only* to consider your evidence. The evidence does matter for an adviser, but further, independent truth-claims *also* matter. This is already sufficient to undermine subjectivism (who claims that *only* ability and evidence-dependent fact matter for moral obligation), which makes this a very compelling argument.

Graham's article moves from that to discussing plausible arguments for subjectivism. He develops interesting counters to them, nevertheless, the nuances of the discussion will not matter so much for us. If someone can hold a subjectivist position that deals with the challenge from OIC above and give a counter for the two arguments, they will be off-the-hook from my arguments below. In general, I think the subjectivist will just deny OIC and the arguments below will also not apply anymore[17]. In those arguments, objectivism will be assumed as part of the presuppositions about when someone is to be blamed or not. And on the existence of criteria about when is it fair to blame someone. If the fairness of acts of blaming is a moral imperative, as I contend it is; and objectivism is true, then the fairness of blaming is also an objective matter. Either a person, when judging other, does so fairly, or not fairly and there is a fact of the matter about which it is.

3. The Biology of Choice and Opportunities

The final stage-setting Section 1 need is connected to the notion of capability. As capabilities are agent's properties and those are instantiated by biological systems, it seems rather important to think about how we come to learn what someone is capable of doing. There are two answers to this question. The first is the dispositional answer. This one is rather simple to answer:

A subject S is justified in believing that some agent Ag is dispositionally capable to do some action A in a set of contexts, if the agent has a track-record of doing A in those contexts; or if a relevantly similar agent Ag has shown this track record and S has no reason to think that Ag* and Ag differ in regarding to the particular context being analyzed.*

This criteria could be improved upon, but it is serviceable. At any rate, this is not the epistemological sense that I think brings difficulties. Rather, the sense I mean is what I will call: "situated capability". This notion of capability is that applied in a particular, predefined context. For instance: you can open doors. But that does not mean that you are capable of opening the heavy, metal door of the example above. Dispositions are not undermined by few counterexamples. You might be able to drive, and drive well, even if you cannot drive after taking a powerful anesthetic. Unfortunately, it does not seem that we have a neat epistemological recipe to learn about situated capabilities. This will be discussed below, but it is important to check some general details of why that might be so difficult.

Vranas defined ability in terms of knowledge, skills and bodily capabilities. Opportunities are those things that allow the agent to put their abilities to good use. Vranas is explicit that psycho-

[17] In fact, these arguments may be taken as evidence against objectivism!

logical properties are not included in the ability part. But I think that it is impossible for them to *not* be included in the opportunity condition for capability. A person's biological make-up might be able to run ten kilometers, they might have done that before, but had they had a bad infection recently, they might not be able to run, say, 3 kilometers. A few weeks later, with good recovery, they might be back to the long-distance running. It is clear that ability was not compromised. All the breathing, rhythm and muscular structure was still there, as well as the know-how of running. What changed is that the body took a hit and had to recover. This is opportunity, not ability, and it matters.

With this in mind, I wish to review the argument put forward by R. SAPOLSKY (2023, ch.3) that we have a nested structure of influences operating under every decision we make. I will not do this to draw his Hard Determinist conclusions. Rather, the claim I wish to make is rather plain: certain factors about our bodies are out of our control, they oscillate based on many, many, variables and those may affect our moral (and otherwise) decisions. The upshot is this: even if libertarian free will is true, and most of our decisions are our own as a non-caused cause; by being possible that we might be affected in the ways that the literature Sapolsky considers, then maybe a particular choice of ours every once in a while was not free. This is compatible with libertarian free will because the fact that external forces might affect our decisions in decisive ways is a plain fact of life. We might be libertarians drugged with special substances that make us temporarily incapable of choosing, for instance. So, I just wish to point out that we may, at times, be in a situation analogous to the drugged libertarian, and on those rare events, we might not be able to choose.

3.1. Sapolsky's argument in a nutshell

I will make a short presentation of the first part of Sapolsky's argument. His earlier book, SAPOLSKY (2017), makes a prolonged case for the same observations, with the important caveat offered by Sapolsky himself (SAPOLSKY, 2023, p. 47 fn.) that his first book suffered from citing some papers that failed the replication crisis in psychology. For brevity, and for the epistemological advantage of only relying on replicated results, I will follow only his (2023, ch.3).

The chapter in question aims at thinking about where does intent come from. The answer, unsurprisingly given that the author is a neurobiologist, has a lot to do with the brain. But not only— it has also much to do with culture and society as well. The review here will be only about the brain, for it suffices to make clear the way in which the opportunity condition may be hard to determine.

3.1.1. Second to minutes before

There is a behavior, why has it occurred? We are talking about humans, and humans are biological creatures. Our brains have something to do with decisions and personality, as the famous case of Phineas Gage illustrates. Brains have substantial control over two crucial systems: neurological and endocrinological systems, or oversimplifying: neurons and hormones.

This section of the book argues for the unconscious, unexpected effects of sensorial stimuli in

moral (or evaluative) decisions[18]. Two main of such cases deserve mentioning in this summary, both are connected to disgust and moral rigidness. This could be inferred weakly from the nature of moral condemnations being frequently associated to disgust, but it is more than words. In a series of experiments, subjects are put in a condition to display moral choices in both uncomfortable situations and disgusting ones. The uncomfortable situation was a bowl of ice-cold water. The disgusting one is an imitation of vomit. The two groups were given the same set of moral situations. Some of them are purity-related moral infractions and others are not (e.g. John rubbed someone's toothbrush in a public toilet's floor vs John risked someone's car with a key). Subjects in both situations had to decide the strength of punishment for the infractions.

Disgusted subjects gave harsher punishments to purity-related infractions than they did to the others. Moreover, their punishment was harsher than the punishment given by the uncomfortable group of subjects (SAPOLSKY, 2023, p. 48).

Still in the connection between disgust and moral rigidness, subjects are asked to make moral judgements about sex and sexuality. One group of subjects was asked to do that in a room with a mild (but noticeable) disgusting smell. The other had a non-disgusting smell in the room. Result: the group in the smelly room leaned more towards conservatism than the other group (SAPOLSKY, 2023, p.47).Who would imagine that a courthouse being built near a smelly area might affect the average decision of judges there?[19]

3.1.2. Minutes to days before

If the previous section presents some ways our brains fail inasmuch as neuronal activations are concerned, in this section we will discuss the problems caused by hormones. The first case to consider is testosterone. Contrary to popular belief, testosterone does not lead to increase in aggression per se. It makes people more prone to defend their social status, especially displaying aggressivity towards those that are *under* them in the hierarchy. Additionally, testosterone affects people who are prone to violence to have a lower threshold to begin violent behavior. Non-violent people seem to be less effected. Finally, it makes us more prone to interpret socially ambiguous interactions (*e.g.* people's faces) as aggressive interactions.

What causes rises in testosterone in people and in its effects? Time of the day, recent activities (fights, discussions, sex, etc). Individual variations. These individual variations are due to genes; fetal, pre and post-natal environments. That means that we have very little control over the levels of testosterone we have when we are about to interpret a social situation in a morally significant way (SAPOLSKY, 2023, p. 52-53).

Similarly, the hormone oxytocin is connected to human bounding. It has multiple social functions and explains much of our social lives. Unfortunately, beyond some beneficial outcomes, it has also surprising effect in in-group favoritism: the psychological disposition to prefer those perceived as in-group than those perceived as out-group. (SAPOLSKY, 2023, p. 54-55). Once more, the effect of this hormone is determined by a myriad set of factors and we can hardly hope to know how it effected a subject in a particular situation in real life.

[18] I will not dwell in the suggested neurological explanations offered by Sapolsky. But each of the examples receives an explanation in the same page in which they are described.

[19] This is me extrapolating from the data. Similar results were tested on legal decisions, but there is more to be done to make a categorical claim that the result is important.

3.1.3. Weeks to years before

This section is about neuroplasticity and bacteria in one's digestive system. The neuroplasticity is attached to how the brain changes over time, building and breaking synapses and/or neurons. It can affect morality in substantial ways. Mental illness and stress can have overarching effects in moral cognition, increasing fear, aggressivity and the like. Likewise, being chronically stressed. Changes for the best can happen also, if the environment is positive (SAPOLSKY, 2023, p. 58).

About the bacteria, we have less to think about in terms of morality directly. But that bacteria in one's digestive system can affect behavioral patterns is a rather surprising fact. Sapolsky lists the following effects observable by switching bacterial makeup between individuals: appetite and food cravings; gene expression patterns in the neurons; disposition towards anxiety and the speed of development of some mental illnesses (SAPOLSKY, 2023, p. 59). If bacteria in one's stomach can alter their behavior, maybe they could do so in a morally relevant context too?

To be very clear, from now on I will be assuming libertarian free will. So, even if bacteria could have some effect, it would usually be out-competed by the agential powers people have. But as a powerful drug might temporarily undermine this agential power, so too a combination of multiple effects like the abovementioned might at times overpower the agential control. The proposition I will be assuming is:

Force-balance (FoB): a strong combination of factors might, on occasion, undermine the agential powers a given libertarian-free-willed agent might have.

If this is the case, the agent loses the opportunity to apply their abilities to the satisfaction of their objectives, inclusive moral ones.

4. Normative Demands, Obligations and Fair Blaming[20]

The next thing I need to defend in my argument is that there is a class of actions that are underexplored, namely, the actions associated to normatively demanding of others. These "normative demands" are second-order actions (actions about actions) concerning how people engage in the practices of morality. This is not the same as, to choose a single term: "normative evaluations". I will store "normative evaluations" to the traditional ethical stance connected to what people (really) ought to do, and normative demands to the acts people engage in when they hold others (and themselves) to some standard, being it a good one or not. Henceforth, I will just say 'demands' and 'evaluations' to mean their normative kinds.

Demands can cause bad or good moral effects on people. For instance, a professor who hire a new PhD student to be their assistant but only demands of them relatively easy cognitive tasks like ("summarize a chapter of an undergrad introduction book") is harming the student by displaying a neglect over their true capacities. In the other direction, if the same professor demands too many high-level tasks ("one knock-down argument against the big players in the literature per week!") this will also cause harm, in this case of overdemanding. In the opposite direction, giving people responsibilities that track their capability is a way of recognizing them as full participants in a given community and to help them to develop. Because of this potential for harm

[20] Professor Macnamara argues against demands in connection to blame. Although I cannot do full justice to her arguments here, one thing I can say is that her view on demands is very different from mine, and most of her objections do not even begin to apply to my work (MACNAMARA, 2013).

and good, I contend that demanding is a kind of action that is passive of moral evaluation. Good demands being called "fair" and bad demands being "unfair". The set of neutral plus fair demands would be called "adequate" demands, whereas "inadequate" is the same as unfair. Presumably, fair demands track obligations, for it would be truly weird if there was no merit in following the best process in ascertaining the adequacy of our demands. Nevertheless, the relation might not be bi-conditional, in fact, there is reason to think that it is not.

I contend that, if what I have argued in Section 2 is correct, then evaluations will correspond to the objectivist perspective taken into account earlier. Whereas, it seems that the fairness of demands are the most plausible candidates for what the subjectivist's intuitions are following. In this interpretation, the Objectivist/Subjectivist debate is actually based on the systematic ambiguity of thinking about what is actually correct versus thinking about what we should fairly demand of people in a given context.

In a nutshell, the issue is the following: We can evaluate agent's demands and conclude that they are bad (or good) in the number of ways that are representative of normal evaluation of actions. But, the differences between evaluations and demands is that of perspective. What is correct for one to demand of others depends on many factors about the demanding agent as well as the demanded subject. This is not supposed to be the case in an objectivist view of evaluation. For this objectivist angle, there are things people are obligated to do, that they are allowed to do, and that they are forbidden to do. These things do not depend on who is evaluating. These judgements are supposed to be (as close as possible to) true-to-the-eyes-of-a-god. On the other hand, subjectivism seems to track the norms about demands quite well, even if it does not do so with obligations. The person in the movie (see Section 2) might be obliged to accuse the murderer for the murder in the living room instead of the suspect, and yet, it might not be fair to demand that they do so because they did what we could reasonably demand of them to do. If I am correct, this shift in the moral camera that captures the Objectivist/Subjectivist debate is a shift from the morality of obligations to the morality of demands.

The existence of such a difference of perspective suggests that the sense in which we are allowed to demand things from other people is tied to their objective obligations only partially: there are also the standards to which we, as judgers, should obey.

This is discussed in COATES & TOGNAZZINI (2013, p. 18-23), where the authors distinguish three ways in which blaming might be adequate. The first are the conditions of blameworthiness; the second are the conditions of *jurisdiction*; and the third are the conditions of procedure. Without prolonging too much on them, the first kind is about what criteria should we look for when deciding if the person is to be blamed. The second— and most relevant here— contends that even if a subject is to be blamed, not all people can blame that person in the same way. Some might be in such a position as to be hypocritical when blaming, and thus losing the fairness of their action. Others might be so far away of the case (in a broad sense of 'far away', not only the geographical one) that their blaming is unfair for not being adequately engaged with some relevant aspects. The third condition is somewhat close to the second, but emphasizes the way in which we pursue blaming. If we administer it with proportionality, or following a good investigation.

I argue here that it is part of the second condition that we should not demand of others more than we could expect of ourselves in the same situation, or more than we can expect of a rele-

vantly similar person in the same situation. How this requirements are to be spelled out is work for the future, yet it is reasonable to say that, within reason, we know with whom to fairly compare cases, and what we should expect of ourselves[21].

Although this works in general for practical purposes, if we wish these intuitive rules of thumb to actually track really existing obligations, we need more epistemological work to be done. Intuitions can fail us in many ways. The trouble in the next section is entirely directed to this asymmetry: we intuitively know when and how to blame people but given that obligations are objective facts; and, if OIC is true, CC is proposition we must be able to affirm or deny in particular cases, then the epistemological requirement for a fair blame will have to be stronger than the intuitions. This, I will argue, is what we cannot deliver.

That said, it would be odd to say that being obligated is not-at-all related to what we are allowed to demand of the subject, thus it is crucial to think about the relations between what we are normatively allowed to demand and what people are obliged to do. Some seemingly plausible rules about it will be relevant for the arguments in Section 5. Let 'agents' to be those people who are demanding of others and 'subjects' to be those who are being demanded:

Obligations and Abstract Demands (OAD): If a given subject has failed to fulfill an obligation, there is, at least in principle, an agent (which might include themselves) who is in position to fairly demand that they had not done so.

Lack of Obligation is the Lack of Demands (LOLD): If a subject is not obliged to Φ, then there is no agent for whom it would be fair to (normatively) demand that the subject should Φ, and/or blaming the subject for not having Φ-ed.

From Obligations to Demands 1 (FOD-1): if a subject is obliged to φ, then it is fair for some agent to demand of them that they should φ, so long as the agent is justified in believing that the subject is capable of φ-ing; and the agent has justified belief that the subject is obliged to φ.

As those prima facie norms are normal moral norms, they should be objectively valid, if objectivism is true. The main reason to hold fair demands, I think, is exactly because, as the principles claim, there is a close relation between a fair demand and objective obligations. This is as it should be, given that, otherwise, morality itself would be inconsistent (and we hope this is not true, although I have very few argument to claim it isn't).

4.1. Fair blame

The foreplay is almost done, so let's set up the last few assumptions needed for the arguments in Section 5. First, I assume libertarianism. And I assume that there is an important sense in which people are responsible for their actions, in the sense of being proprietary, answerable causal beginners that acted for motives connected to the action, and who could have acted otherwise. Therefore, I am presupposing that there are legitimate ways to blame and punish people for what they have done. This being said, I am assuming that libertarianism is true, but not that we

[21] I am certain that the reader is skeptical so here is some remarks. We might always be allowed to expect other subjects to (i) act in conformity to instrumental rationality. (ii) act in accordance with their role's duties (e.g. to treat people if one is a doctor); and (iii) to perform in accordance with objectively valid conditions (e.g. in accordance with the true moral system, or in ways that are avoiding clear failures).

know that it is true, anymore than that we know it right now. That is, although people will be uncaused causes of their own behavior most of times, we, as outsiders evaluating people and making demands on them will have no more access to this fact than we currently do. There is no "free-will-do-metre".

To assume, however, that blame is sometimes due is not to claim that it cannot be misdirected; being applied to people who do not deserve (so much of) it. Hence, granting that blame is morally acceptable or even necessary is not the whole story, we must discuss under which conditions that blame can be fairly attributed. If it cannot be fairly attributed then it must either be only a descriptive tool to capture a given inclination of resented people, or it is an incorrect, unfair response and thus immoral. To be clear, I have the following definition of fair blame:

Fair Blaming (FB): A judgment of blame is done fairly, if the agent doing it is: i) *well-informed*, meaning that the judgement respects a conscientious amount of morally relevant information about the judged situation; ii) *well-reasoned*, meaning that the arguments are sound, bear relevant premises and are not defeated by the full-body of morally relevant information taken into consideration, or that ought to be taken into consideration; iii) *adequately universal*, meaning that the individual, the arguments and principles and all relevant moral information lead to a judgment of blame that would apply equally to any person in the same situation. Finally, iv) *proportionality*: the judgment of blame must accept some degree of proportionality between the badness of the action, the amount of responsibility the subject had and the amount of blame we allocate to that subject.

First, *well-informedness* requires conscientious appeal to the available morally relevant properties. This means both taking in consideration all the evidence already available and applying adequate standards for further investigation. This is obviously vague and highly contextual. But, although the details matter greatly in general, for this argument the reader is invited to fulfill it with their favorite theory of epistemic inquiry[22]. Regardless, it is hard to imagine a world in which a poor investigation would yield a fair blame. Additionally, if no good account of what an adequate investigation into blameworthiness is possible, my main argument will run even better, so for the sake of strengthening the position of my adversaries, I shall concede that some such investigation is possible. *Well-reasonedness* does not need much defense. Obviously, if an agent tries to blame some subject based on poor reasoning, the judgement will not satisfy FOD-1 and, hence, will not be a demand that tracks a duty. *Adequate universality*, is the condition of non-ad-hoc-ness and this is equally obviously necessary. If I just make up standards for each person by fiat this arbitrariness is bound to bring forth injustice. And finally, *proportionality* is not very important for my argument, but something like it must be added if we wish to have an adequately complete account of fair blaming. In the end, I think these defining conditions do offer us a good candidate of a definition for Fair Blaming. If not, they are at least necessary conditions, and this much is what I hope to argue from.

So, if i)-iv) are at least necessary conditions, what is the problem I wish to advance? The problem is that i)-iii) are not plausibly met when it comes to token people. I will take some time to present this concept.

[22] Mine would be that of GOLDBERG's (2018) and his Hybrid account. Although the theory faces some push-back from SIMION (2024) and PEELS (2019).

4.2. Token and Type people

Abstraction is a crucial part of hypothetical reasoning, but many judgements of blame are actually conducted against a living, flesh-and-bones person. The difference between hypothetical people like the best possible Brazilian, or Jack who must decide if he pulls the lever in a Trolley Problem-type, or the average German; and concrete people, is that in the first case, their doings— and the morally relevant properties of their situations— are all established by stipulation or abstraction, not so when it comes to the concrete ones out there. Let's call the abstract people: *type people*; and the concrete kind: *token people*.

The great difference between those two kinds of people is the situations in which they enact moral actions, type-people do so in thought experiments whereas token people do so in the real world. More precisely, although we know all about hypothetical scenarios, and we can wish out of existence some unfriendly complexities; in real life, all actual events are imparting in the moral action, as the results discussed in Section 3 above suggest. This is not to say that there are no irrelevant properties for moral consideration, sure there are. Usually, it is an irrelevant fact about a murder attempt whether or not it happened near a quartz rock. Therefore, I grant that there are cases of properties that are morally irrelevant and that those can be ignored when thinking about type or token people's moral character or blameworthiness.

The problem is quite different, in fact. It concerns the range of morally relevant properties being investigated and found out by both moral psychology and philosophy. This number is growing every year with new findings about how humans actually think morally and what they take into consideration when judging to act. This is not to say that the morality of a given decision is itself to be found in the lab behavior of monkeys or in brain activity. Rather, the issue is about what people are actually capable of doing: the degree to which the moral capability of people is affected by those ever-growing numbers of known morally relevant properties. The reason this matters is because OIC makes capability into a robust moral property: one whose presence or absence changes the fairness of a judgement by itself.

4.3. Token OIC

OIC applies to moral duties, and those apply to people. If there are two kinds of people, then it seems fair to consider whether we have two kinds of OIC. It is probable that there are because the fairness of an agent's demands over a subject that they should φ, depends on whether the subject had ability and opportunities and whether the agent knew that[23]; and this issue can depend on whether the context is hypothetical or empirical. If it is hypothetical, then all the capacities and opportunities can be determined by fiat. If it was empirical, then we have to make a case that the particular person had, on that particular case, the ability and the opportunity. Given that 'ability' are the skills, knowledge and agential bodily functions, it is usually easier to determine if the person was able. However, the opportunity depends on the person having the right conditions at her disposal to enact their ability. But, given that the variability of those properties that impact our behavior can be very wide, and in some extremes they are like a powerful drug in their incapacitating powers, it is possible that the subject — even having the ability— might be biologically constrained, thus lacking the opportunity to impart that ability

[23] Given LOLD and OIC, if there is a lack of capability, then there can be no fair demands to the effect that Φ.

into getting a duty done. To repeat: this is compatible with libertarianism, for libertarians only would contend that *most of the time* we have free will, not that *all of the time* we have such. The comparison with the drug is important. If we have free will *all of the time* then a drugged person is fully responsible. But they are not. Hence, we do not. Equally, I hold, for the specific conditions in which FoB (Section 3) applies, which I am assuming are quite rare.

Given that the issue I wish to bring out is on the fairness of demands, we will focus on the agent (the person blaming), the subject's capability being plain or not (having the opportunity or not) is not very relevant. It would be if we had a 'free-will-do-metre', but we do not. Absent that, the problem remains: How could an agent demanding that a subject should do φ be justified in believing that the subject had (at that particular frame of time) the opportunity to φ? Without that justification, the demand will not track obligations failing FOD-1; and will not satisfy the condition i) of FB. Hence, without that justification, the demand for φ-ing would not be fair, and is unlikely to be connected to a true obligation the subject has. This is the crux of the argument, so the reader is advised to go back and reacquaint with the principles and definitions mentioned.

Then, the crux of the matter is that if we, as agents, cannot determine if the subject is in one of those rare moments in which multiple biological variables came together to undermine the subject's opportunity to impart their abilities into fulfilling their duties, we would fail FB and FOD-1 above. The reason being that, if OIC is true, capability is a robust morally relevant property: one whose presence of absence makes a difference by itself into blame or not-blame.

We have two ways of ascertaining moral blame, one is strictly speaking abstract and hypothetical, the other is highly empirical. Both depend on the same principles and definitions of fairness, and thus both should be equally valuable in guiding our moral judgements. But, naturally, the empirical version— the one directed at blaming token people— is much harder to do, and a lot more consequential.

5. The Epistemological Arguments Against Token Blame

5.1. The argument from full investigation

In this section, I advance the most direct argument against the possibility of guaranteeing the conditions of FB for token people. I will call "the capability of a token person to fulfill their duties" their 'token capability'; and "subject's capability" when it goes for both type and token people. In the next section, I will advance a more nuanced version, yet, both start with the same premises, which will be simply numbered; and when the premises are unique to one of the two arguments it will be named with numbers and lowercase letters. This first argument has two versions. It moves like this:

1) We can only blame subjects fairly if we conscientiously investigate the morally relevant properties regarding their situation under consideration (intuitively plausible).

2) The subjects' capability is a morally relevant property regarding the situation, of such a kind and relevance that all conscientious investigations must justifiably ascertain their presence or absence (from OIC and capability being robust).

C1) We can only blame token people fairly if we justifiably take into consideration their token

capabilities (follows from 1,2).

3a) The number of known conditions for capability has increased as research in moral psychology, moral cognition, neuroscience, social psychology, philosophy, etc. has developed (SAPOLSKY (2017, 2023), briefly summarized in Section 3 above[24]).

4) The list of moral variables is, now, so vast and difficult to account for; and the research of their impact is so contextual, that we cannot justifiedly determine whether their co-occurrence in a given context would be sufficient to undermine subject's capability. (Section 3, SAPOLSKY, 2017; SAPOLSKY, 2023); Section 1: CC's requirement fulfillment is underdetermined)

5a) Additionally, we cannot determine if the situation's description is ever complete and include all the opportunity-relevant properties. (given that we are learning more and more, and that we lack a general law for predictions, we cannot affirm that our current knowledge is exhaustive).

6a) If we cannot know that a given description of opportunity-relevant properties is complete, nor when the overall co-occurrence of those properties will satisfy FoB, then we cannot be sure when judging the presence of token capabilities.

C2) If we are never sure whether token capabilities are present, we always fail necessary conditions for token-Fair Blaming.

C2 leads to the rejection of Fair Blaming because we probably fail FB's Criterium i), for there is no way to decide if all the relevant information was considered, hence if we do actually consider them it was out of luck. And we fail criterium iii) because by not knowing if we have all the relevant properties we cannot be sure which cases are morally analogous to which and, hence, reasoned universality is fatally compromised.

I find this argument particularly convincing against condition i). To sum up the argument, I claimed as follows. Fair blame requires conscientious analysis of the morally relevant properties. A crucial one is the capability people have to act according to duty. This property is crucial because of OIC; but it is also an empirical matter whether a given token person has the opportunity to actually fulfill their duties. It is not only an empirical matter as it is also an impossible practical task, given that it requires understanding factors that go back to evolutionary developments and cultural implications, among others. Therefore, we are never able to perform an adequate analysis of the morally relevant properties in token people, 'opportunity' will allude us. This is not to say that the subject had, or didn't have, the opportunity. It is only to say that the agent is not justified in their belief one way or the other; and because fairness requires this justification according to FOD-1 and FB, the demand is not fair— the act of blaming is not fair. But because this situation is pervasive, no acts of blame could be fair.

5.1.1. A second version of the same argument

Much the same argument can be run in a slightly different way. This time the argument is based on a different attempt to select morally relevant properties. It departs from a different approach to fair blaming, namely: take universality and proportionality as the sources of your judgement.

Method: We start with a past case whose blame is really plausible and infer from that case that the current one is also to be blamed. This strategy has a juridical ring to it, but it promises to

[24] See TIBERIUS (2014) for a book-length introductory review.

avoid the problem above if only by assuming that, at times, we have blamed correctly. I do not want to draw a complete skeptical scenario from the outset, so I will grant the hypothesis.

Hence, the strategy we are contemplating is the following: Stage 1- get a comparison case (C-Case) from which to judge your actual case (A-Case). Stage 2- highlight similarities and differences between C-Case and A-Case. Stage 3- punish the A-Case in accordance with the C-Case plus or minus the aggravating or exculpating factors given by Stage 2-. Plausibly, if these three stages are done appropriately, then we can grant a satisfaction to FB necessary conditions, even if they are slightly weaker in this case. So, under this background, the new version of the argument goes like this:

1) We can only blame subjects fairly if we conscientiously investigate the morally relevant properties regarding their situation under consideration.

2) The subjects' capability is a morally relevant property regarding the situation of such a kind and relevance that all conscientious investigations must justifiably ascertain its presence or absence.

C1) We can only blame token people fairly if we justifiably take into consideration their token capabilities.

3b) We have a C-Case from which to compare and get the information about our A-Case's token capabilities. (by hypothesis)

4) The list of moral variables is, now, so vast and difficult to account for; and the research of their impact is so contextual, that we cannot determine whether their co-occurrence in a given context would not be sufficient to undermine subject's capability.

5b) The C-Case is only useful in blaming token people if we can conscientiously assess the differences between the C-Case's token capabilities and the A-Case's token capabilities. (plausible assumption about analogical reasoning)

6b) if 4 is true, we cannot empirically guarantee that if C-Case was fairly blamed, then A-Case was fairly blamed, for the effect of C-Case's morally relevant properties might have been context-dependent, or some exclusively A-Case set of properties might defeat the inference that held for C-Case.

Hence, 6b concludes the same as C2 before. This argument, actually, requires that we can separate well the properties of the cases between those that matter and those that do not. But, because in token people we cannot, the analogy will not offer justification, rather, it will presuppose that all morally relevant properties (for a conscientious investigation) are the same between both cases. But the very issue at hand is to suggest that we cannot guarantee that! This new strategy, however, opens up to a second line of criticism that attacks the assumption of analogousness.

5.2. The Argument from the Lack of Analogies

The second kind of argument takes pains to show that, if we are considering token people, we cannot just assume that two cases will be analogous. I will defend later that there are properties with what I will call "idiosyncratic moral effects". These are defined as properties whose moral significance varies substantially – they can in one case be able to allow for fair blaming and in

another to allow for full exculpation – depending on with which other properties they are co-instantiated; and it requires that consequence to be different both in quality and in intensity.

1) We can only blame subjects fairly if we conscientiously investigate the morally relevant properties regarding their situation under consideration.

2) The subject's capability is a morally relevant property regarding the situation of such a kind and relevance that all conscientious investigations must justifiably ascertain its presence or absence.

C1) We can only blame token people fairly if we justifiably take into consideration their token capabilities.

3c) To argue from analogy between different token cases (or from type people to token people) we must guarantee that the analogy holds and agent's capabilities are the same in both cases (at least inasmuch as the presence or absence is the same).

4) The list of moral variables is, now, so vast and difficult to account for; and the research of their impact is so contextual, that we cannot determine whether their co-occurrence in a given context would not be sufficient to undermine subject's capability.

Two options for analogies to hold.
Option 1:

— 5c)[25] An analogy holds for morality if the differences between cases, even if they are morally relevant, are not morally relevant for the current situation (foundational principle of analogical reasoning).

—— 6c) If properties can have idiosyncratic moral effects, there is no class of reference from which to justifiably infer whether a given moral property was morally relevant to another situation or not (follows from 5c and the definition of idiosyncratic moral effects)

——— Consequently, the analogy is not justified.

Option 2:

— 7c) An analogy holds for morality if the differences between cases, even if they are morally relevant, are not as impactful in the current situation as the similarities.

—— 8c) If properties can have idiosyncratic moral effects, we cannot rule out the hypothesis that the differing properties might have more impact in one case than in the other such as to trump the impact of the similarities (due to 4 and the definition of idiosyncratic moral effects).

———Consequently, the analogy is not justified.

9c) if either 6c) or 8c) are true; and if properties can have idiosyncratic moral effects, then we have no justified source to affirm the analogy between the fairness of demands due to the supposed parity between two seemingly equivalent cases. (We are gratuitously assuming that the differences are not (sufficiently) relevant i.e. begging the question).

10c) Properties can have idiosyncratic moral effects.

C3) We cannot obtain fair judgements of blame by analogy with either other token people or with type people.

[25] '—' is being used, as is standard in Natural Deduction, to mark a hypothesis and the hierarchy of dependence. As the hypothetical are not included in a single body of argument, but bifurcate in two different ones.

C4 Hence, if both C2) and C3), then we cannot satisfy Fair Blame at all.

This argument rejected the possibility of plausible universality, undermining FB's condition iii). But, by doing that, it rejected an important strategy to highlight morally relevant properties— analogy— which makes the argument also undermine i) by making it explicit that analogies are incapable of offering us all the needed information about morally relevant properties. The above argument requires two supporting pieces of evidence. One to argue for premise 10c, and another to argue for the importance of properties having idiosyncratic moral effects.

To defend 10c), although this is not a discussion to be settled with a single argument, I find that the hypothesis is highly plausible. It just states that in different cases, the same property has different moral effects to the extent that the very same property might make one person to be blamed in a situation and to be exculpated in another due only to the contribution of these properties to the other morally relevant ones. Moreover, it requires that consequence to be different both in quality and in intensity. Given the variability of human moral behavior, the idea that a set of three properties like "short-temper, obsessive motivation and high conceptual intelligence" could not lead to different outcomes in different people is rather unlikely. Additionally, a plausible example might suggest that they are not even that hard to come by.

Imagine a subject who don't have their right arm. Not having a right arm may be irrelevant when considering the fairness of a demand over that subject for donating money for charity. Nevertheless, it might be very relevant when deciding whether the subject was a murderer in a case in which forensics guarantee that the gun was held in the suspect's right hand. If this example is plausible, the property "having a right arm" has idiosyncratic moral effects in the sense I am using the terms. It absolutely does not matter in one case, and absolutely exculpates in the other.

Alright, if properties with idiosyncratic moral effects can exist, why do they matter for premises 6c,8c, and 9c? The reason is quite simple. If a given property might, when added to others, bring about *surprising*, novel moral effects, then by choosing the two supposedly analogous cases based on their similarity in moral properties alone will not suffice, given that this idiosyncrasy emerges from the overall combination of properties—moral and otherwise. On the other hand, stipulating that the cases are similar enough will just beg the question. This, once more, only applies to token people. When it comes to type people, we can stipulate that the cases are analogous given that we can determine the absence of idiosyncrasies in the cases.

The last thing that I must argue is that some of those novel moral effects may affect capability. But this is rather obvious. Imagine two men: Johny and Najib, who are both, at around the same time, in a very similar situation. Both men had received orders from their spouses to hold on to a plate of sweets and candies while they prepare everything for their children's party. There are not many candies and if even one is taken the beautiful decoration their spouses worked hard to get will be undone. Both are sweet-lovers and those sweets are their favorites. Once more, both men had had super stressful weeks, and they are prone to sugar eating when wishing to compensate for stress. However, Johny's mother had a substantial trauma when we was still a fetus and because of this, he acquired a decreased self-control relatively to Najib, Johny usually can deal with that quite well. However, it so happens that when three factors get together, Johnny's self-control runs out. The conditions are: stressful weeks, favorite sweets, and low blood sugar caused by

more-than-average time without eating[26]. Consequently, Johnny, but not Najib, eats one sweet and get their spouse's angry response for it. Now, the property "having had a stressful week" prevented Johny, but not Najib from fulfilling a duty. Had they been instantiated in slightly different conditions, both men would had been analogous relatively to this particular scenario. But in the next scenario, we would not be able to know. This is why the existence of properties with idiosyncratic moral effects is a problem for these kinds of token people analogies. The only issue remaining is whether this sort of example is compatible with the dialectical assumption of libertarian free will. For the particular case that I am making here, it doesn't matter, for the objective is to show that idiosyncratic moral effects can affect capability by undermining opportunity. But if there is an inconsistency with libertarian free will, then there might exist more implications for the argument above. I cannot pursue the issue here, for it would require a precise presentation of free will accounts.

To sum up, this argument, if successful, undermines our capacity to learn about morally relevant properties in situations of blaming by asking us to justify our analogies regarding whether or not a given token agent exhibits an ability.

6. Recapitulating and exploring exits

So, here is the eagle-eye point of view. We started discussing blame as an objective thing grounded on the way people are and in what they do. I further added that obligations are also objective things grounded on the way the world is. I discussed some reasons to think that, even if libertarian free will is true, there might be some combination of co-occurrent properties that, on occasion, could block the opportunity to act according to one's own free will. I argued that this is analogous to the case of a strong drug making people do things they would not ordinarily do. I then presented a distinction between token and type people. I moved to argue that there is one morally relevant property i.e. a subject's capability that must be known to be present or not, and, still about which we get to learn very little when it comes to token people. Because conscientious investigation was one condition for fair blaming, and because those arguments show that this is impossible to do, we concluded that fair blaming was impossible.

A conclusion nobody wants, I guess. At any rate, I don't. So, this section explores how to avoid it, that is, I will take the arguments in the previous section to be *reductio ad absurdum* of the set of propositions we brought in the beginning, with the exclusion of libertarian free will. That view plays only the didactical role of showing that the discussion here is very loosely connected to free will. The set of assumptions that lead to the problems are these: Realism about blame, objectivism about obligation, moral normativity over demands, epistemic fallibility in the sense of not being able to know all the propositions we could possibly know in the relevant time, OIC (capability), and the empirical discoverability of new morally relevant properties. Idiosyncratic moral effects are not necessary for the first argument, but are for the second. Finally, of course, there was the two conditions for FB (i) and iii), and FOD-1. We might try to deny each of them, and although this could be done in a myriad ways, I will discuss just the most obvious and broad consequences.

[26] For these kinds of effects in real life, Sapolsky's 2023 book offer a very useful compilation and explanation. The fetal effect is derived from the really observed consequences of the Dutch Hunger: See ROSEBOOM *et al* (2011) for a review of these effects. For space's sake I preferred to just add the small, grounded fantasy above than doing a dive into those papers here.

6.1. First option, undermining OIC

If OIC is false, then the individual's ability to fulfill their duties becomes less important, and the rest of the morally relevant properties may be more tractable epistemically speaking. The OIC-sceptic must, however, still contend with the intuitive idea that people's capability matters to their duties. Hence, this strategy requires a substitutive theory, not merely a rejection of OIC. For different reasons, KING (2019, ch.4) explores such a view. I think her account is unsuccessful to face this challenge, and it fails on some internal grounds, but it would be a way out[27].

6.2. Second option: Embracing

It is interesting— and somewhat ironic— to think about the implications of OIC to avoid the conundrum above. Given that the knowledge of capability is impossible, and, hence, we have a practical impossibility, if OIC is true, then the duty to blame fairly may be wavered. In which case, we are in the clear to blame unfairly, as we both cannot avoid to blame (for an argument to this effect SHER (2006 ch.1) and cannot hope to blame fairly. That is a consistent way out of my argument, if only a bit unorthodox. In other words, the application of OIC to the act of judgement makes it the case that we might not be obliged to blame fairly, because we are incapable of doing so. Hence, blaming may just not be a thing we can ever do morally. This denies the normativity of demands above. Some people like PEREBOOM (2001) and SAPOLSKY (2023) might wish to just agree with this move, after all, they already do not believe that there is such a thing as fair blame.

6.3. Third option: change the definition

Maybe there is no requirement of conscientiousness in the moral inquiry required for fair blaming. This may well be, but most of my arguments above only needed an argued position about the effect of one property of the agent— namely: 'opportunity'— consequently, to avoid the issue definitionally might be harder than one might think. This option combined with rejecting OIC may seem more promising. At any rate, it seems hard to reject that demanding— and blaming as a subset of that— require some adequate evidential standard, and this standard clearly seems to need to account for all the properties that might, on their own, undermine blame: the 'robust properties' as I called them.

This is the case of opportunity. If there was opportunity, the person might be fairly blamed. If there wasn't, they cannot. The fact that we do not know either way may be taken to be negligible because most times people can act according to their duties. But I don't think this is so easy. We know that there are excusing circumstances at multiple times— sometimes even fully exculpating ones— so the odds that FoB happens is not *so low* as to be negligible by fiat. When we do the empirical investigation, we frequently see those factors playing in, FoB conditions are not *ad-hoc*. That is different from considering a claim for the lack of opportunity given the intervention from 4D alien Spacecraft, or any some such scenario. In this new scenario, but not in FoB, we have *no evidence at all* of the existence of these aliens. This point needs to receive more attention if this argument is to truly fly, but it is *prima facie* plausible to maintain that all robust moral properties should be adequately investigated.

[27] A forceful attempt against OIC is offered by J. GOLDWATER (2020).

6.4. Fourth option: Acceptance

I have been dealing with the hypothesis that moral judgement justification must be somewhat strong, that is because most people take that morality is something we are very good at; maybe we can even do it *a priori* based on intuitions like Moore seem to have defended. Therefore, the epistemic conditions have high standards. But maybe we are bad at moral judgements, in this case, we might just be high fallibilists about morality and just go with it. In which case, the definition of Fair Blaming might be: the argument to the effect of blame was plausible. This has many meta-ethical implications. WEDGWOOD (2007) does argue for a fallibilist view on our moral epistemology, but given the above set-up, the fallibilism will be very strong. We will be mostly adding a huge luck element in our investigations to the effect that, if I investigated all I could in the best way available to me, it is likely that the fairness of blame will track obligations adequately.

6.5. Fifth option: blame is directly related to demands, not to obligations

Maybe the problem with this whole project is the metaphysical realist background. I tend to think that this is correct. Perhaps the idea that people cause bad things and then get to be blamed by them, given that not doing bad things was an obligation they really had is what does not mix well with the empirical nature of our actions. The suggestions here would be the following. Perhaps we should decouple blame from failing an obligation and link it more directly to the fairness of demanding that the person had not done that. Then, blame would require the justified belief that the subject of blame has causally participated in the coming-to-be of the bad outcome (in the relevant way that exclude spurious causal claims) and the fairness of the demand that the person had not done so. Now, because the fairness of the demand only require justified beliefs, and the absence of opportunity might not lead to unjustified beliefs about obligations, only untruthful ones, it would follow that the argument above is dismantled. The mere possibility that the subject might not be capable of delivering on their duty would not undermine the fairness of an agent's demands about it. By taking blame one step back from obligation, we can hold blameworthiness even if in the end the person was not truly too obligated, due to their incapacity. It will all depend on the fairness of the demands. This exit denies the mentioned principle in Section 4.1 called LOLD. The counter intuitive side of it is that we can be properly blamed by things that we, in the end, were not truly obligated to do (or for not failing to do things we were not obliged to do). Indeed, this is a big bullet to bite.

Conclusion

This paper argues for the unexpected tension between some standard positions. OIC, objectivity about blame and obligation, and the empirical nature of human capability. If the arguments are correct, something about it must be rejected. This tension, I think, was neglected for long because a crucial difference, with robust methodological implications, was not observed, or not observed enough: the difference between token and type people. Taking that difference seriously puts us into the path of facing the epistemic challenges any person trying to judge the blameworthiness of another must face. This paper is not a defense of the claim that there is no fair blaming, rather it is a defense of the option five (Subsection 6.5), or it would had been had I had the space for that. In any case, these arguments presented in section V are the only arguments I

know of that reject the possibility of fair blaming, and consequently, large parts of moral responsibility, without touching upon debates about free will, except in the lightest of ways. This, if not anything else, may show that they merit philosophical attention.

Bibliographic References

BASSFORD, A. D. 2022. Ought implies can or could have. *The Review of Metaphysics*, [*S.l.*], v. 75, n. 4, p. 779-807.

BERGLUND, A. 2005. *From conceivability to possibility*: an essay in modal epistemology. 197 p. Aarhus, Denmark. Doctoral dissertation. Filosofi och lingvistik. Aarhus University.

COATES, D. J.; TOGNAZZINI, N. A. 2013. The contours of blame. In: COATES, D. J. (Ed.); TOGNAZZINI, N. A. (Ed.). *Blame*: its nature and norms. Oxford: Oxford University Press.

FRANKFURT, H. G. 1969. Moral responsibility and the principle of alternative possibilities. *Journal of Philosophy*, New York, v. 66, n. 23, p. 829-839.

GOLDBERG, S. 2018. *To the best of our knowledge*: social expectations and epistemic normativity. Oxford: Oxford University Press.

GOLDWATER, J. 2020. Six Arguments Against 'Ought Implies Can'. *Southwest Philosophy Review*, [*S.l.*], v. 36, n. 1, p. 45-54.

GREENE, J. 2013. *Moral tribes:* emotion, reason, and the gap between us and them. New York: Penguin.

GRAHAM, P. A. 2010. In defense of objectivism about moral obligation. *Ethics*, Chicago, v. 121, n. 1, 88-115, Oct 2010.

HIERONYMI, P. 2020. *Freedom, resentment, and the metaphysics of morals*. Princeton: Princeton University Press.

KING, A. 2019. *What we ought and what we can*. London: Routledge.

MACNAMARA, C. 2013. Taking demands out of blame. In: COATES, D. J. (Ed.); TOGNAZZINI, N. A. (Ed.). *Blame*: its nature and norms. Oxford: Oxford University Press.

PARFIT, D. 1971. Personal identity. *Philosophical Review*, New York, v. 80, n. 1, p. 3-27.

PEELS, R. 2019. The social dimension of responsible belief: response to Sanford Goldberg. *Journal of Philosophical Research,* Virginia, v. 44, p. 79-88. https://doi.org/10.5840/jpr201944150.

PEREBOOM, D. 2001. *Living without free will*. Cambridge: Cambridge University Press.

ROSEBOOM T. J.; PAINTER R. C.; VAN ABEELEN A. F.; VEENENDAAL M. V.; DE ROOIJ S. R. 2011. Hungry in the womb: what are the consequences?: lessons from the Dutch famine. *Maturitas*, Ireland, v. 70, n. 2, p. 141-145, Oct 2011. https://doi: 10.1016/j.maturitas.

SAPOLSKY, R. M. 2023. *Determined: Life without free will*. New York: Random House.

SAPOLSKY, R. M. 2017. *Behave: The bestselling exploration of why humans behave as they do*. New York, Random House.

SCANLON, T. M. 2013. Interpreting blame. In: COATES, D. J. (Ed.); TOGNAZZINI, N. A. (Ed.). *Blame*: its nature and norms. Oxford: Oxford University Press.

SCANLON, T. M. 2008. *Moral Dimensions:* permissibility, meaning, blame. Cambridge, Mass: Harvard University Press.

SHER, G. 2005. *In praise of blame*. Oxford: Oxford University Press.

SIMION, M. 2024. *Resistance to evidence*. Cambridge: Cambridge University Press.

SMITH, A. 2013. Moral blame and moral protest. In: COATES, D. J. (Ed.); TOGNAZZINI, N. A. (Ed.). *Blame*: its nature and norms. Oxford: Oxford University Press.

STRAWSON, P. F. 2008. *Freedom and resentment and other essays*. London: Routledge.

TIBERIUS, V. 2014. *Moral psychology*: a contemporary introduction. London: Routledge.

VRANAS, P. B. 2024. Beyond ought-implies-can: impersonal obligatoriness implies historical contingency. *Journal of Ethics and Social Philosophy*, v. *29*, n. 1, p. 29-61.

VRANAS, P. B. 2018. "Ought" implies "can" but does not imply "must": an asymmetry between becoming infeasible and becoming overridden. *Philosophical Review*, New York, v. 127, n. 4, p. 487-514.

VRANAS, P. B. 2018. I ought, therefore I can obey. *Philosophers' Imprint,* Michigan, v. 18, n. 1, p. 1-36, Jan 2018.

VRANAS, P. B. 2007. I ought, therefore I can. *Philosophical studies*, Netherlands, v. 136, p. 167-216, Jul 2007.

WEDGWOOD, R. 2018. The unity of normativity. In: STAR, D. (Ed.). *Oxford handbook of reasons and normativity*. Oxford: Oxford University Press.

WEDGWOOD, R. 2016. Objective and subjective 'ought'. In: CHARLOW, N. (Ed.); CHRISMAN, M. (Ed.). *Deontic modality*. New York: Oxford University Press.

WEDGWOOD, R. 2007. *The nature of normativity*. Oxford: Oxford University Press.

WILLIAMS, B. 1973. *The problems of the self*. Cambridge: Cambridge University Press.

Para além da (ir)racionalidade: *différance*, fantasia e uma outra educação

Beyond (ir)rationality: différance, fantasy and another education

Diogo Bogéa
Universidade do Estado do Rio de Janeiro (UERJ)
diogobogeaa@hotmail.com

Abstract: Tentamos nesse artigo repensar a questão da racionalidade e da irracionalidade sociais a partir de referenciais desconstrucionistas e psicanalíticos. Iniciamos definindo racionalidade social como projeto comum orientado por um fundamento suposto absoluto que pré-determina sentidos e finalidades supremos. Trazemos então as noções de *différance*, cálculo e responsabilidade de Derrida para apontar a impossibilidade dos projetos racionalistas e encaminhar possibilidades ético-políticas (des)orientadas pela alteridade, pela singularidade e pela incalculabilidade. Em seguida, abordamos a questão da adesão em massa às chamadas *fake news* como exemplo de fenômeno geralmente atribuído à "irracionalidade social", mas que, com nossos referenciais desconstrucionistas articulados à noção psicanalítica de *fantasia*, ganha novas cores. Por fim, trazemos teorias educacionais/curriculares contemporâneas que, inspiradas por referenciais teóricos desconstrucionistas e psicanalíticos ressituam os processos educacionais em torno da diferença, da alteridade, da singularidade e da incalculabilidade – questionando, portanto, os projetos educacionais racionalistas que se estruturam em torno de ideais de comunidade e calculabilidade.

Keywords: *différance*; fake news; fantasia; responsabilidade; singularidade; teoria curricular.

Resumo: In this article, we attempt to rethink the issue of social rationality and irrationality through deconstructionist and psychoanalytic frameworks. We begin by defining social rationality as a common project guided by a presumed absolute foundation that predetermines supreme meanings and ends. We then introduce Derrida's notions of *différance*, calculation, and responsibility to highlight the impossibility of rationalist projects and to outline ethical-political possibilities (dis)oriented by otherness, singularity, and incalculability. Next, we address the mass adherence to so-called *fake news* as an example of a phenomenon typically attributed to "social irrationality," which, when reexamined through our deconstructionist references articulated with the psychoanalytic notion of *fantasy*, takes on new contours. Finally, we present contemporary educational and curricular theories that, inspired by deconstructionist and psychoanalytic approaches, reframe educational processes around difference, otherness, singularity, and incalculability—thus questioning rationalist educational projects structured around ideals of community and calculability.

Palavras-chave: *différance*; fake news; fantasy; responsibility; singularity; curricular theory.

Recebido em 7 de julho de 2025. Aceito em 14 de outubro de 2025.

Considerações iniciais

Há um curta de Godard (1962) chamado *O Novo Mundo* (*Le nouveau monde*) cuja trama retrata uma explosão nuclear que gera como efeito colateral o emburrecimento e a estupidificação generalizados da espécie humana. Apenas o protagonista, que durante a catástrofe dormia, escapa dos efeitos colaterais e tem de viver na mais aguda solidão como último lúcido num mundo em que todos se tornaram irremediavelmente estúpidos. Nesse "novo mundo", todos parecem agir mecanicamente, são acometidos por um grave empobrecimento da linguagem, as palavras já não têm significado, eles dizem e desdizem na mesma frase, repetem gestos estranhos automaticamente, tomam remédios compulsivamente. Trata-se de uma incisiva crítica à massificação que produz uma vida automatizada, sem profundidade e sem sentido. Após um sono profundo, aquele que desperta está radicalmente só, transitando em meio a uma multidão de zumbis. Enquanto isso, os jornais, escritos eles mesmos por idiotas, afirmam sistematicamente que tudo está bem e que a explosão nuclear não teve qualquer efeito. Na última cena do curta, o protagonista escreve em seu caderno: "O mundo novo começou e um milagre me salvou. Mas eu também posso estar contaminado pela terrível mecanicidade... a morte da lógica".

Embora compreendamos a beleza e a atualidade do curta de Godard, colocaremos em questão, nesse artigo, seu diagnóstico para a "terrível mecanicidade": a "morte da lógica". Tomando como referência alguns elementos do pensamento derridiano e da teoria psicanalítica, a relação entre lógica/racionalidade e mecanicidade (passível de ser lida como emburrecimento ou estupidificação generalizada) se torna menos clara, mais ambígua e, talvez, surpreendentemente íntima.

Traremos em nosso auxílio, ainda, reflexões contemporâneas sobre Educação e Currículo. A Educação nos parece um campo privilegiado para a discussão proposta por esse dossiê, pois é vista como o meio, por excelência, de *racionalização* social. De estabelecimento de uma ordem social racional na qual todos os sujeitos "educados" possam exercer sua plena "autonomia" no sentido kantiano. Daí a interminável disputa em torno do Currículo: o que se deve transmitir às pessoas para que realizem a plenitude da sua essência como sujeitos racionais? Pensadores contemporâneos do Currículo, justamente a partir de referenciais teóricos derridianos e psicanalíticos, vêm problematizando esse sentido fortemente racionalista atribuído aos processos educacionais.

1. Do que estamos falando quando falamos em racionalidade e irracionalidade?

Embora já tenhamos adiantado alguns pontos da nossa investigação nas considerações iniciais pedimos ao leitor que nos acompanhe numa espécie de passo atrás: talvez antes ainda de oferecer respostas à questão sobre a racionalidade e a irracionalidade sociais devêssemos nos perguntar o que estamos chamando aqui de racionalidade e irracionalidade.

Racionalidade remete a *ratio*, a tradução latina de *legein*, raiz de *logos*. A etimologia relaciona *legein* a *reunir, juntar, colher, agrupar* (CASSIN et al., 2014, p. 581). Como diz o *Dicionário de Intraduzíveis* coordenado por Barbara Cassin: "O que é intraduzível" na raiz de *logos*, "é a unidade subjacente à ideia de 'reunir', uma série de conceitos e operações — matemáticas, racionais, discursivas, linguísticas — que, a partir do latim, são expressas por palavras que não guardam qualquer relação entre si"(CASSIN et al., 2014, p. 581)[1]. O dicionário oferece como exemplo o

[1] Tradução minha do inglês: "What is untranslatable here, paradigmatically, is the unity beneath the idea of 'gathering together,' a series of concepts and operations—mathematical, rational, discursive, linguistic—that, starting with Latin, are expressed by words that bear no relationship to one another."

latim *ratio* e *oratio*, que guardam afinidade com diferentes significações que emanam da palavra *logos*. *Ratio* diz "computar", "calcular", "contar" e, por extensão, "pensar". *Oratio*, em íntima relação com *oris*, a boca, remete a "palavra", "fala", "discurso". Dessa ambiguidade deriva aquela da palavra *contar* em português, francês e inglês: do latim *computare*, *contar* é tanto calcular e lidar com números, quanto contar uma história, narrar, dizer.

No mais famoso trecho do princípio de sua *Política*, Aristóteles apresenta o logos como determinação essencial do humano ao mesmo tempo em que articula essa racionalidade humana à racionalidade social. Acompanhemos o texto:

> É evidente que o homem é um animal mais político [*politikòn*] do que as abelhas ou qualquer outro ser gregário. A natureza, como se afirma frequentemente, não faz nada em vão, e o homem é o único animal que tem o dom da palavra [*lógon*]. E mesmo que a mera voz sirva para nada mais do que uma indicação de prazer ou de dor, e seja encontrada em outros animais (uma vez que a natureza deles inclui apenas a percepção de prazer e de dor, a relação entre elas e não mais que isso), o poder da palavra [*lógos*] tende a expor o conveniente e o inconveniente, assim como o justo e o injusto. Essa é uma característica do ser humano, o único a ter noção do bem e do mal, da justiça e da injustiça. E é a associação de seres que têm uma opinião comum [*koinonia*] acerca desses assuntos que faz uma família [*oikian*] ou uma cidade [*polis*] (ARISTÓTELES, 2004, p. 146).

Aqui se enuncia uma noção que fará história no pensamento ocidental: o "próprio" do humano, o que torna o humano diferente dos demais animais é o *logos*, esse poder ordenador que *reúne* a dispersividade das expressões sensíveis de dor e prazer em *conceitos* de bem e mal, justo e injusto. O *logos* reúne a multiplicidade de sensações e opiniões particulares na unidade de um conceito universalmente compreensível pela comum-unidade dos seres racionais. Assim, o poder ordenador e reunidor do *logos* agrega também os humanos em comunidades. A chave desse princípio de reunião humana é "o dom da palavra", pois o *logos* reúne as palavras em um discurso articulado. Seguindo o fio dessa ordem que compõe a articulação das palavras é possível produzir discursos bem fundamentados que se distinguem da mera opinião.

A *reunião* – como a "colheita" à qual o *legein* também está intimamente ligado – se faz segundo um critério previamente estabelecido de seleção. Esse princípio prévio que se estabelece como critério é o próprio *arkhé*, o *fundamento* que preside a reunião da multiplicidade existente de entes, experiências e opiniões na unidade do Ser, do conceito e do discurso bem fundamentado. Seguir o fio do raciocínio lógico é se aproximar da descoberta desse fundamento supremo previamente dado. Por isso o sábio platônico é aquele que se aproxima da contemplação dos *eide*, os modelos proporcionais eternos e perfeitos que presidem a composição das formas do mundo sensível.

O princípio pré-estabelecido se dá também como *finalidade* suprema da vida individual e coletiva. Buscar os primeiros princípios é também buscar as finalidades últimas. Por isso mesmo Aristóteles faz do "primeiro movente" a *causa final* suprema. O "primeiro movente", que move "sem ser movido", que é "substância eterna e ato", "move como o que é amado" [buscado, desejado], isto é, como finalidade suprema do movimento, "enquanto todas as outras coisas movem sendo movidas" (ARISTÓTELES, 2002, p. 563).

Racionalidade social, segundo os termos até aqui colocados, se refere, portanto, à vida no âmbito de um projeto comum a todos, com base em um fundamento bem estabelecido tendo em vista uma finalidade suprema pré-determinada por esse mesmo fundamento.

Ecos desse projeto racionalista platônico e aristotélico seguirão reverberando nos projetos iluministas de racionalização total da vida. O melhor exemplo talvez seja a noção kantiana de

autonomia que faz coincidir a suprema liberdade do sujeito "universalmente legislador" à incondicional obediência à "lei que ele próprio se dá" (KANT, 2017, p. 374). O imperativo categórico "age segundo uma máxima que possa valer simultaneamente como lei universal!" (KANT, 2017, p. 35) deve se sobrepor a qualquer inclinação particular como dever supremo de todos os seres racionais. Não há contradição, para Kant, entre liberdade e obediência nesse nível, pois esta "lei moral" funda-se sobre a "autonomia de sua vontade, como uma vontade livre que, de acordo com suas leis universais, necessariamente tem de ao mesmo tempo poder concordar cum aquilo ao qual deve submeter-se" (KANT, 2016, p. 211-212). Mas esse dever pessoal só pode se consumar verdadeiramente no interior de um Estado regido por leis racionalmente estabelecidas, leis com tendência à universalidade: "o Direito é (...) o conjunto das condições sob as quais o arbítrio de cada um pode conciliar-se com o arbítrio de outrem segundo uma lei universal da liberdade" (KANT, 2017, p. 231). Por isso a finalidade suprema da Natureza para o gênero humano é "uma constituição civil perfeitamente justa" (KANT, [*s.d*], p. 9) no interior de uma "constituição estatal interiormente perfeita" (KANT, [*s.d*], p. 15).

O meio necessário para a realização dessa plena racionalização social é a educação. Nem Aristóteles nem Kant o ignoravam. Aristóteles afirma que

> Como não há senão um fim comum a todo o Estado, só deve haver uma mesma educação para todos os súditos. Ela deve ser feita não em particular, como hoje, quando cada um cuida de seus filhos, que educa segundo sua fantasia e conforme lhe agrada; ela deve ser feita em público. Tudo o que é comum deve ter exercícios comuns. É preciso, ademais, que todo cidadão se convença de que ninguém é de si mesmo, mas todos pertencem ao Estado, de que cada um é parte e que, portanto, o governo de cada parte deve naturalmente ter como modelo o governo do todo (ARISTÓTELES, 2006, p. 78).

Kant, em seu pequeno tratado *Sobre a Pedagogia*, compreende que "o homem não pode se tornar um verdadeiro homem senão pela educação. Ele é aquilo que a educação dele faz" (KANT, 1999, p. 15). Kant lamenta a tamanha diversidade nos modos de vida humanos: "Na verdade, quanta diversidade no modo de viver ocorre entre os homens! Entre eles não pode acontecer uma uniformidade de vida, a não ser na medida em que ajam segundo os mesmos princípios" (KANT, 1999, p. 17). A Educação deve uniformizar a diversidade humana segundo o princípio supremo da moralidade – o *dever* – e a finalidade suprema que é "a Ideia de uma república perfeita, governada segundo as leis da justiça" (KANT, 1999, p. 17). Para isso, uma educação para a *autonomia* – isto é, para a obediência a leis racionalmente estabelecidas – deve ser providenciada desde cedo:

> A selvageria consiste na independência de qualquer lei. A disciplina submete o homem às leis da humanidade e começa a fazê-lo sentir a força das próprias leis. Mas isso deve acontecer bem cedo. Assim, as crianças são mandadas cedo à escola, não para que aí aprendam alguma coisa, mas para que aí se acostumem a ficar sentadas tranquilamente e a obedecer pontualmente àquilo que lhes é mandado, a fim de que no futuro elas não sigam de fato e imediatamente cada um de seus caprichos (KANT, 1999, p. 13).

Ao longo dos séculos XIX e XX, com a concretização dos projetos de escolarização em massa, as primeiras teorias de Currículo têm também essa forte vocação racionalizadora. Surgidas nos Estados Unidos no início do século XX, as teorias eficientistas, progressivistas e aquelas que propunham uma síntese entre ambas enfatizam

> o caráter prescritivo do currículo, visto como planejamento das atividades da escola realizado segundo critérios objetivos e científicos. (...) Admitindo-se o caráter científico de sua elaboração, os insucessos são, com frequência, descritos como problemas de implementação e recaem sobre as escolas e os docentes (LOPES; MACEDO, 2011, p. 26).

As prescrições curriculares, baseadas em princípios pré-estabelecidos cientificamente devem

ser implementadas tendo-se em vista a finalidade última da educação: a formação de trabalhadores qualificados e cidadãos organicamente adaptados à racionalidade social.

Encaminharemos a partir daqui algumas problematizações ao ideal de racionalidade social até aqui exposto. Nos valeremos, para tanto, de alguns elementos do pensamento de Jacques Derrida e da teoria psicanalítica.

2. O problema da reunião

Um dos motes pelo qual o pensamento de Derrida é mais conhecido é a crítica ao chamado *logocentrismo*. O que nos parece estar, decisivamente, em jogo na crítica derridiana ao logocentrismo é justamente uma crítica do *logos* como princípio de *reunião*. Retomando a linguística estrutural de Saussure, Derrida encaminha outra compreensão da linguagem como não mais fundada por uma ordem reunidora, mas como a-fundada numa diferencialidade abissal. Uma das mais precisas descrições dessa releitura derridiana de Saussure se encontra no texto *La différance*, de 1968. A partir dos princípios saussurianos da arbitrariedade do signo e da relacionalidade diferencial que confere sentido ao signo conforme sua posição numa certa cadeia ou sistema de significação, Derrida extrai uma lógica da *diferença*:

> O conceito de significado não é nunca, em si mesmo, presente, numa· presença auto-suficiente que não remeteria senão para si mesma. Todo o conceito está por direito, inscrito numa cadeia ou num sistema no interior do qual remete para o outro, para os outros conceitos, pelo jogo sistemático das diferenças. Em semelhante jogo, a diferença não é mais, portanto, um conceito, mas a possibilidade da conceitualidade, do processo e dos sistemas conceituais em geral. Por esta mesma razão, a diferença [*différance*], que não é um conceito, não é uma simples palavra, ou seja, aquilo que representamos como sendo a unidade calma e presente, auto-referente, de um conceito e de uma fonia (DERRIDA, 1991, p. 42).

Différance, aqui, não diz respeito à diferença entre identidades plenamente constituídas, mas ao jogo da diferencialidade geral que "produz" e possibilita toda significação e identificação, ao mesmo tempo em que impede e impossibilita toda suposição de plenitude para qualquer significação ou identidade. A *différance* é a "'origem' não-plena, não-simples" (DERRIDA, 1991, p. 43) de toda significação. Ela se insinua nas frestas de espaçamento e temporização que fazem com que toda significação ou identidade só possa vir a ser em relação com *outras*, guardando a marca de *outras* significações passadas e em função de significações futuras:

> A diferença [*différance]* é o que faz com que o movimento da significação não seja possível a não ser que cada elemento dito "presente" , que aparece sobre a cena da presença, se relacione com outra coisa que não ele mesmo, guardando em si a marca do elemento passado e deixando-se já moldar pela marca da sua relação com o elemento futuro, relacionando-se o rastro menos com aquilo a que se chama presente do que àquilo a que se chama passado e constituindo aquilo a que chamamos presente por intermédio dessa relação mesma com o que não é ele próprio: absolutamente não ele próprio, ou seja, nem mesmo um passado ou um futuro como presentes modificados. É necessário que um intervalo o separe do que não é ele para que ele seja ele mesmo, mas esse intervalo que o constitui em presente deve, no mesmo lance, dividir o presente em si mesmo, cindindo assim, como o presente, tudo o que a partir dele se pode pensar, ou seja, todo o ente na nessa língua metafísica, particularmente a substância e o sujeito: Esse intervalo constituindo-se, dividindo-se dinamicamente, é aquilo a que podemos chamar espaçamento, devir-espaço do tempo ou devir-tempo do espaço (temporização). E é a esta constituição do presente como síntese 'originária': e irredutivelmente não-simples, e portanto, stricto senso, não-originária, de marcas, de rastros de retenções e pretensões (...) que eu proponho que se chame arqui-escrita, arqui-rastro ou diferença [*différance*]. Esta (é) (simultaneamente) espaçamento (e) temporização (DERRIDA, 1991, p. 45).

Aqui se encontra também a descrição de outro dos motes pelos quais o pensamento de Derrida é conhecido: a desconstrução da *metafísica da presença*. É que o jogo da diferencialidade, o mesmo que produz toda significação, impede ao mesmo tempo que uma significação se encontre consigo mesma, se reaproprie de si mesma, constituindo uma plena presença-a-si. Partindo

dessas noções, Derrida evitará se referir a tal dinâmica da diferencialidade como *linguagem*, mas sim como *escritura* e também não mais falará em *significante* ou *significado*, mas em *rastro*. Essencialmente diferencial, o rastro é aquilo que, nem presente nem ausente, se dá como condição de (im)possibilidade de toda dinâmica de significação e ressignificação, aparecimento e encobrimento, identificação e diferenciação. O (im) aqui posto antes de possibilidade não se presta a ser um mero artifício retórico, mas procura expressar o jogo de uma diferencialidade abissal que possibilita toda presentificação ao mesmo tempo em que impossibilita qualquer plenitude de presença.

A atenção a essa dinâmica da diferencialidade tende a fazer com que se encare com profunda suspeita e desconfiança todo discurso que pretende reunir um "todos nós" qualquer, todo discurso que propõe se referir à reunião de "todos nós" numa identidade plenamente constituída, todo discurso que pretenda estabelecer as origens, os sentidos e as finalidades legítimas para "todos nós", ainda que – e talvez mais ainda quando – esse discurso pretende falar pelo bem de "todos nós". Como muito bem observa Paulo Cesar Duque-Estrada:

> Por mais nobre que seja, uma igualdade que reúne todos em um "nós", por exemplo, "nós, os humanos", é sempre uma igualdade afirmada, postulada, instituída. Dito de outro modo, ela se estabelece como um ato performático, e, nesse sentido, não pode jamais ser entendida como alguma coisa que já existisse por si mesma, em sua presença disponível e comum a todos. Em termos desconstrucionistas, por mais nobre que possa ser este "nós" – ao qual cada indivíduo deve ser devidamente restituído – ele não impede nunca a validade, a necessidade, a pertinência, e mesmo a urgência de se perguntar "nós quem?", "quem diz nós?", "de que lugar se diz nós?", "com que critérios ou pressupostos?", "com vistas a que se diz nós?" (DUQUE-ESTRADA, 2004, p. 43).

A impossibilidade da presença plena faz com que a discursividade tenha que assumir – embora muito frequentemente sem que verdadeiramente se assuma isso – uma certa performatividade. Como não há nem referente nem significado plenamente presentes, toda dinâmica de referência se faz atravessada por mediações, seleções, edições, ocultamentos, obliquidades, opacidades. Tal dinâmica não está sob o comando de nenhum sujeito ou instituição específicos, pois todo sujeito ou instituição já são eles mesmos efeitos dessa mesma dinâmica. Trata-se de uma performatividade intrínseca à própria dinâmica diferencial da escritura que, impedindo o acesso pleno, claro e transparente a qualquer acontecimento, faz com que a referência só seja possível a partir de uma interpretatividade:

> Uma interpretação faz o que ela diz, enquanto ela pretende simplesmente enunciar, mostrar e ensinar; de fato, ela produz, ela é já de uma certa maneira performativa. De modo naturalmente não dito, não confessado, não declarado, faz-se passar um dizer do acontecimento, um dizer que faz o acontecimento por um dizer do acontecimento (DERRIDA, 2012, p. 237).

Assim, dizer "todos nós, humanos" ou "todos nós" de qualquer maneira é performativamente pré-supor e assim efetivamente tentar estabelecer uma *identidade* única e homogênea para "todos nós". No projeto kantiano, portanto, de estabelecer uma comum-unidade dos sujeitos racionais no interior de uma constituição civil e estatal perfeitamente racional e justa, permanece impensado o ato performativo de fundação de um tal Estado, ato sempre necessariamente *externo* ao regime legal que ele mesmo funda e legitima. Eis o problema da soberania:

> A operação de fundar, inaugurar, justificar o direito, fazer a lei, consistiria num golpe de força, numa violência performativa e portanto interpretativa que, nela mesma, não é nem justa nem injusta, e que nenhuma justiça, nenhum direito prévio e anteriormente fundador, nenhuma fundação pré-existente, por definição, poderia nem garantir nem invalidar. Nenhum discurso justificador pode, nem deve, assegurar o papel de metalinguagem com relação à performatividade da linguagem instituinte ou à sua interpretação dominante. O discurso encontra ali seu limite: nele mesmo, e, seu próprio poder performativo (DERRIDA, 2010, p. 24-25).

O "nós", pré-suposto e imposto pelo ato performativo que supõe falar "por todos nós", "em nome de todos nós", encarna um movimento de apagamento da *différance*, das alteridades, das singularidades. "Toda e qualquer identidade se configura no âmbito de uma ordem – 'como eu, como nós' – do mesmo". E o *mesmo*, "é sempre autorreferencial, apropriador, protecionista, exclusivista: na ordem do *mesmo* sacrifica-se o *outro*" (DUQUE-ESTRADA, 2020, p. 49). Perguntado por Maurizio Ferraris sobre a repetida aparição em suas obras da suspeita em relação à noção de "família" encarecida por André Gide, Derrida responde:

> Eu não sou da família significa, em geral, "eu não me defino com base no meu pertencimento à família", ou à sociedade civil, ou ao Estado; eu não me defino com base nas formas elementares de parentesco. Mas também significa, mais figurativamente, que eu não sou parte de nenhum grupo, que eu não me identifico com uma comunidade linguística, uma comunidade nacional, um partido político, ou com qualquer grupo ou 'panelinha' de qualquer tipo, com qualquer escola filosófica ou literária. "Eu não sou da família" significa: eu não me considero "um de vocês", "não conte comigo como um de vocês", eu quero manter minha liberdade, sempre: esta é, para mim, a condição não apenas para ser singular e 'outro', mas também para entrar em relação com a singularidade e a alteridade dos outros. Quando alguém é da família, não apenas se perde na horda, mas perde os outros também; os outros se tornam simplesmente lugares, funções de família, ou lugares ou funções na totalidade orgânica que constitui um grupo, escola, nação ou comunidade de sujeitos falantes da mesma língua (DERRIDA; FERRARIS, 2001, p. 27).[2]

O que está em jogo para Derrida na desconstrução do *logocentrismo* é, portanto, a desconstrução do *logos* como poder de *reunião*. Desconstrução que se encaminha para além da lamentação kantiana pela existência de tamanha diversidade entre os humanos. Mas também para além da "celebração" das diversidades compreendidas como identidades plenamente constituídas em si e por si mesmas com especificidades "próprias" e modos de vida "apropriados". A desconstrução se faz em nome de uma atenção à *singularidade* absoluta de todo e qualquer *outro*. Singularidade irrepresentável, inapreensível por uma lógica da identidade:

> toda vez que um rastro é transformado em algo, em o que quer que seja "como tal"; toda vez que um rastro se encerra no paradigma da presença – perdendo, assim, a sua intrínseca heterogeneidade – em favor de uma – suposta e institucionalizada – homogeneidade e autoidentidade; toda vez que isso acontece, é a "realidade" mesma das coisas ou, em termos mais derridianos, a sua singularidade, o seu caráter único, que é excluída. Essa é a violência fundamental: a exclusão da singularidade. O que se encontra fundamentalmente em jogo aqui é, como disse, a resistência, a resposta, o enfrentamento mesmo dessa violência; em outras palavras, é a tentativa de se fazer justiça à singularidade em seu caráter de rastro; libertá-la de sua redução à verdade objetiva, à norma do "como tal", ao paradigma da presença (DUQUE-ESTRADA, 2020, p. 39).

3. O problema da pré-determinação

O *logos* não se refere jamais a uma reunião aleatória, mas a uma reunião segundo um princípio e uma finalidade pré-estabelecidos. Remete a um saber prévio quanto aos princípios que estabelecem os critérios e as finalidades da reunião. A *colheita*, um dos sentidos de *legein*, nos fornece um bom exemplo: quem sai para colher morangos não deve misturar em seu balaio gravetos, pedras, insetos ou frutas de qualquer outro tipo. Isso não é algo que se decide em meio à colheita. É o saber prévio que orienta e possibilita a própria colheita.

[2] Tradução minha do inglês: "I'm not one of the family means, in general, 'I do not define myself on the basis of my belonging to the family', or to civil society, or to the state; I do not define myself on the basis of elementary forms of kinship. But it also means, more figuratively, that I am not part of any group, that I do not identify myself with a linguistic community, a national community, a political party, or with any group or clique whatsoever, with any philosophical or literary school. 'I am not one of the family' means: do not consider me 'one of you', 'don't count me in', I want to keep my freedom, always: this, for me, is the condition not only for being singular and other, but also for entering into relation with the singularity and alterity of others. When someone is one of the family, not only does he lose himself in the herd, but he loses the others as well; the others become simply places, family functions, or places or functions in the organic totality that constitutes a group, school, nation or community of subjects speaking the same language".

Esse saber prévio que pré-determina o sentido dos entes é justamente o que Heidegger chama de *cálculo*. No ensaio *A época das imagens de mundo*, ao tratar da matematização das ciências naturais entre a modernidade e a contemporaneidade, Heidegger explicita por que a matemática é o campo paradigmático do cálculo:

> *Ta mathemata* significa para os gregos aquilo que o homem, na consideração do ente e na lida com as coisas, conhece à partida: dos corpos o corpóreo, das plantas o vegetal, dos animais o animal, do homem o humano. A este já conhecido, isto é, ao matemático, pertencem também, para além do que foi referido, os números. Quando encontramos três maçãs na mesa, reconhecemos que são três delas. Mas o número três, a tríade, já o conhecemos. Tal quer dizer: o número é algo matemático. É só porque os números apresentam como que o mais patente sempre-já-conhecido, e, deste modo, o que é mais conhecido entre o matemático, que o matemático foi logo reservado para a nomeação do que é próprio dos números. Mas de modo nenhum a essência do matemático é determinada pelo numérico (HEIDEGGER, 2002, p. 100).

Leitor de Heidegger, Derrida tomará para si esse sentido de *cálculo* como saber prévio que pré-determina o sentido de todo outro que vem ao encontro – fazendo com que o outro não nos chegue verdadeiramente como outro, mas já enquadrado numa lógica do mesmo. Numa lógica calculadora, o outro é sempre reduzido àquilo mesmo que eu suponho saber que ele é. Essa compreensão do cálculo como saber prévio que pré-determina o sentido de toda lida com o outro fica clara em *Força de Lei*, quando Derrida explora a tensão entre o Direito e a Justiça. "O direito não é a justiça. O direito é o elemento do cálculo, é justo que haja um direito, mas a justiça é incalculável, ela exige que se calcule o incalculável" (DERRIDA, 2010, p. 30).

Como se vê na própria frase, Derrida não é contra o cálculo:

> É preciso o cálculo e jamais fui contra o cálculo, você o sabe, a reticência condescendente, a altivez "heideggeriana". Mas o cálculo é o cálculo. E se falo tão frequentemente do incalculável ou do indecidível, não é por simples gosto pelo jogo ou para neutralizar a decisão, ao contrário: creio que não há nem responsabilidade nem decisão ético-política que não deva passar pela prova do incalculável ou do indecidível. Não haveria senão cálculo, programa, causalidade, na melhor das possibilidades, "imperativo hipotético" (DERRIDA, 2018, p. 170).

O que está em questão, para Derrida, é repensar o cálculo para além de um programa previamente estabelecido e bem conhecido. É pensar o cálculo como aquilo que ocorre entre o parâmetro previamente já sabido e um radical não-saber que sobrevém diante da singularidade dos outros e dos acontecimentos.

Essa é a dinâmica que Derrida procura exprimir nas três aporias que encerram a primeira parte de *Força de Lei*. A aporia da *epokhé* da regra traz a experiência da "decisão justa". Para ser considerada justa, uma decisão deve estar esteada em algum tipo de saber prévio, ou seria meramente aleatória. Porém, ainda para ser considerada justa, uma decisão deve poder se desprender do previamente estabelecido para considerar as especificidades de cada caso.

> Em suma, para que uma decisão seja justa e responsável, é preciso que, em seu momento próprio, se houver um, ela seja ao mesmo tempo regrada e sem regra, conservadora da lei e suficientemente destruidora ou suspensiva da lei para dever reinventá-la pelo menos na reafirmação e na confirmação nova e livre de seu princípio. Cada caso é um caso, cada decisão é diferente e requer uma interpretação absolutamente única, que nenhuma regra existente ou codificada pode nem deve absolutamente garantir. Pelo menos, se ela garante de modo seguro, então o juiz é uma máquina de calcular (DERRIDA, 2010, p. 45).

A segunda aporia é a *assombração do indecidível*. "Indecidível é a experiência daquilo que, estranho, heterogêneo à ordem do calculável e da regra, *deve*, entretanto, (...) entregar-se à decisão impossível, levando em conta o direito e a regra" (DERRIDA, 2010, p. 46). O jogo do *rastro* e da *différance* instauram uma dinâmica assombrada pelo indecidível. Não há um sujeito plenamente constituído, presente a si, plenamente racional e consciente que possa, a partir do *domínio* de um saber prévio bem estabelecido, *decidir*. Não há identidades plenas – entre os entes ou significa-

ções – que possam aparecer *enquanto tal*, decididamente, de modo tal que o sujeito não possa errar ao decidir sobre o que são e como lidar com elas. O indecidível permanece assombrando toda lógica da decidibilidade "como um fantasma". Essa "fantasmaticidade desconstrói do interior toda garantia de presença, toda certeza ou toda pretensa criteriologia que nos garanta a justiça de uma decisão" (DERRIDA, 2010, p. 48).

A terceira aporia diz respeito à "urgência que barra o horizonte do saber" (DERRIDA, 2010, p. 51). É preciso decidir. Decide-se. É essa a lei da responsabilidade: *responde-se*, de alguma maneira, querendo mais, ou menos, sabendo mais, ou menos, com palavras, gestos, olhares ou silêncio. Responde-se. O outro, essa alteridade abissal, toda outra, que há em qualquer outro, nos interpela, nos chama a responder, exige respostas nossas. A decisão "é uma loucura":

> Uma loucura, pois tal decisão é, ao mesmo tempo, superativa e sofrida, conservando algo de passivo ou de inconsciente, como se aquele que decide só tivesse a liberdade de se deixar afetar por sua própria decisão e como se ela lhe viesse do outro. As consequências de tal heteronomia parecem temíveis, mas seria injusto eludir sua necessidade. Mesmo que o tempo e a prudência, a paciência do saber e o domínio das condições fossem, por hipótese, ilimitadas, a decisão seria estruturalmente finita, por mais tarde que chegue, decisão de urgência e de precipitação, agindo na noite do não-saber e da não-regra. Não da ausência de regra e de saber, mas de uma re-instituição da regra que, por definição, não é precedida de nenhum saber e de nenhuma garantia como tal (DERRIDA, 2010, p. 52).

O grande interesse ético-político da desconstrução, aliás, é a tentativa de fazer justiça à singularidade do outro – inclusive de qualquer "si mesmo", que é sempre também "outro". Singularidade incalculável, inapresentável, inapreensível, irrepresentável, mas que por isso mesmo exige e impulsiona toda uma dinâmica do cálculo, da apresentação, da apreensão, da apropriação, da representação. "Aquilo que desafia a antecipação, a reapropriação, o cálculo — qualquer forma de predeterminação — é a *singularidade*" (DERRIDA; FERRARIS, 2001, p. 21). Supor saber quem o outro é e como lidar com ele de acordo com um programa previamente definido é reduzir a alteridade à mesmidade, a singularidade a um padrão de identidade.

> Essa é a violência fundamental: a exclusão da singularidade. O que se encontra fundamentalmente em jogo aqui é (...) a resistência, a resposta, o enfrentamento mesmo dessa violência; em outras palavras, é a tentativa de fazer justiça à singularidade em seu caráter de rastro; libertá-la de sua redução à verdade objetiva, à norma do 'como tal', ao paradigma da presença (DUQUE-ESTRADA, 2020, p. 39).

O que está em jogo, portanto, em todas essas articulações é a *desconstrução* dos ideais logocêntricos de *racionalidade social*. O sonho do puro cálculo é assombrado pelos espectros do incalculável, a autonomia é assombrada pela heternonomia, o dever universal é assombrado pela invenção particular, o saber prévio é assombrado pelo não saber, a racionalidade é assombrada pela loucura. Subvertendo os ideais kantianos, a lei não garante a justiça, a autonomia é inseparável da heteronomia, o dever é (des)orientado pela singularidade – e não orientado para a universalidade.

4. A questão das fake news

O curta de Godard que citamos nas considerações iniciais fala de um mundo pós-apocalíptico no qual os habitantes da Terra são acometidos por uma terrível estupidez. Godard – ou o personagem principal do seu curta – descreve tal situação como "terrível mecanicidade... a morte da lógica". Difícil evitar a associação entre o mundo pós-apocalíptico retratado por Godard e o fenômeno contemporâneo de disseminação das chamadas *fake news*. Resistiremos à tentação de enumerar os mais desconcertantes exemplos – dos quais cada leitor terá um bom número em mente.

Em dissertação defendida em 2020, João Pedro Martins realiza um mapeamento de obras que tratam do tema "pós-verdade", termo que ganhou notoriedade após ser escolhido como "palavra do ano" do *Dicionário Oxford* em 2016. Segundo Martins, "em quase todas as publicações se aponta, inicialmente, uma preocupação crescente em relação à queda da noção de verdade" (MARTINS, 2020, p. 44-45). Essa preocupação se faz acompanhar de uma outra relativa à "natureza bizarra dos diversos argumentos e perspectivas que ganham, no dia a dia, a grande mídia e são veiculadas por vezes como absurdos", mas que "parecem ganhar a opinião e a confiança de milhões de pessoas" (MARTINS, 2020, p. 45).

Algumas das descrições recorrentes associadas ao diagnóstico de "declínio da verdade", como aponta Martins são: a "substituição da razão pela emoção", a diminuição do "valor da verdade", o "relativismo pernicioso" (MARTINS, 2020, p. 49) e uma suposta ampliação da receptividade do público para a mentira – em público (MARTINS, 2020, p. 52). Outro elemento recorrente do diagnóstico é a "crise dos valores da Revolução Científica e do Iluminismo" associada à ascensão de vertentes teóricas ou movimentos sociais tais como "Relativismo, pós-modernidade (ou pós-modernismo), Nova Esquerda (*New Left*) e Desconstrução" (MARTINS, 2020, p. 53). Estes últimos são frequentemente apontados como herdeiros da apostasia nietzschiana em relação à fé na verdade ao afirmar que "não há fatos, apenas interpretações" ("Contra o positivismo, que fica no fenômeno 'só há fatos', eu diria: não, justamente não há fatos, só interpretações" (NIETZSCHE, 2008, p. 260).

Em todos os exemplos analisados vemos mais, ou menos explicitamente uma proposta de retorno à razão e à verdade, resgate da razão e da verdade embora, paradoxalmente, como aponta Martins, a maioria deles procure deixar claro que não acredita num "passado idílico no qual se falaria a verdade sempre" (MARTINS, 2020, p. 51). Ao contrapor o reino da pós-verdade ao reino da racionalidade, os autores analisados, explícita ou implicitamente, atribuem o curioso fenômeno da disseminação de – e adesão em massa às – *fake news*, a algum tipo de *irracionalismo social.*

Traremos em nosso auxílio alguns elementos da teoria psicanalítica a fim de encaminhar outra compreensão para a questão contemporânea das *fake news*. Talvez valha destacar desde o princípio que rejeitaremos a própria noção de "irracionalismo" – seja como diagnóstico, seja como suposto projeto teórico "pós-moderno" – por compreender que "irracional" é um modo de adjetivação que só fará algum sentido se aceitarmos os parâmetros, padrões e critérios estabelecidos pelos ideais racionalistas de base platônica e aristotélica. Com uma abordagem desconstrucionista e psicanalítica, por exemplo, tais padrões caem e, com eles, o próprio sentido da noção de "irracionalismo".

Temos a impressão de que para compreender a adesão em massa às *fake news*, para além das espectralidades derridianas precisamos recorrer a uma outra lógica e uma outra fantasmalidade: a *lógica do fantasma* psicanalítica, que introduz um elemento fundamental no jogo da *différance*: a *fantasia*. A espectralidade derridiana remete à dinâmica diferencial que faz com que um sentido só possa "aparecer" através dos muitos "outros" passados e futuros (no tempo e no espaço) que possibilitam sua aparição, ao mesmo tempo em que impossibilitam sua presença plena. A *différance* confere à realidade um caráter espectral, sendo o espectro "um indecidível (nem isto nem aquilo, nem vivo nem morto, nem corpo nem alma, nem dentro nem fora, nem passado nem presente, sempre *milieu* / meio, ponto de partida, no entanto, para qualquer decisão)" (SO-

LIS, 2014, p. 91). Já a lógica psicanalítica do fantasma, como veremos, introduz nessa dinâmica o *desejo* de um gozo absoluto absolutamente impossível, que nos parece fundamental para compreendermos o fenômeno contemporâneo da adesão em massa às *fake news*.

Para Lacan, desde que entramos na linguagem – e esse "desde" não faz referência a um tempo específico, mas a uma condição incontornável – o real está, para nós, fundamentalmente perdido. O "acesso" ao real só se faz através do *simbólico*, da rede de significações que compõem esse "Outro": a dinâmica da linguagem. "Este real, para apreendê-lo, não temos outros meios – em todos os planos e não somente no do conhecimento – a não ser por intermédio do simbólico" (LACAN, 1985, p. 128). Mas, ao mesmo tempo, há um tipo de encontro imediato com esse real, uma certa experiência que, então certamente não é de "apreensão", mas de encontro traumático com o nada, o vazio, o buraco do real. Há então uma "revelação do real naquilo que tem de menos penetrável, do real sem nenhuma mediação possível, do real derradeiro, do objeto essencial que não é mais um objeto, porém este algo diante do que todas as palavras estacam e todas as categorias fracassam" (LACAN, 1985, p. 209).

Esse objeto essencial que não é mais um objeto é o que Lacan chamara de *objeto a*, o objeto primeiro, o objeto "real" do desejo da espécie e do sujeito. Esse objeto, no entanto, é fundamentalmente perdido. Falta desde sempre – "desde" que a linguagem nos exila do real. "No mundo humano, a estrutura como ponto de partida da organização objetal é a falta de objeto" (LACAN, 1995, p. 55). Essa falta funda a dinâmica do *desejo* como busca de reencontro e reapropriação do *objeto a* – que é também busca de preenchimento e sutura do buraco real.

> Uma nostalgia liga o sujeito ao objeto perdido, através da qual se exerce todo o esforço da busca. Ela marca a redescoberta do signo de uma repetição impossível, já que, precisamente, este não é o mesmo objeto, não poderia sê-lo. A primazia dessa dialética coloca, no centro da relação sujeito-objeto, uma tensão fundamental, que faz com que o que é procurado não seja procurado da mesma forma que o que será encontrado. É através da busca de uma satisfação passada e ultrapassada que o novo objeto é procurado, e que é encontrado e apreendido noutra parte que não no ponto onde se o procura. Existe aí uma distância fundamental, introduzida pelo elemento essencialmente conflitual incluído em toda busca do objeto (LACAN, 1995, p. 13).

Em torno dessa falta fundamental se inscreve um pequeno enxame de significantes que marcam o sujeito como um "traço unário", uma marca significante a partir da qual, em sua articulação com uma certa rede dinâmica de significantes do campo do Outro (o simbólico, a linguagem) se estrutura a fantasia do sujeito. "Sujeito" aqui não deve nos enganar com a suposição de que se trata de um sujeito substancial, consciente, racional. O "próprio" do sujeito é ser "outro", diferir de "si mesmo". Tudo o que lhe é "próprio" vem do Outro – do simbólico como rede dinâmica de significantes. "O corpo próprio é, originalmente, esse lugar do Outro, enquanto é aí que se inscreve a marca enquanto significante" (LACAN, 2008, p. 379). Por isso Lacan falará sempre no Sujeito "dividido", "barrado" do acesso ao seu "real" enquanto tal.

Em torno da falta real fundamental, isto é, em torno do *objeto a* fundamentalmente perdido, e em articulação com o traço significante unário impresso nessa borda, proliferarão fantasias simbólicas que prometem o preenchimento, a completude, a unificação, o gozo. A *fantasia* compõe "cenas", "roteiros", "histórias". Assim, "há sempre uma cena em que o sujeito é apresentado no roteiro sob formas diferentemente mascaradas, na qual ele é implicado em imagens diversificadas" (LACAN, 1999, p. 422). Fantasia é a cena que envolve o sujeito e o situa numa certa relação com o objeto "pequeno a":

> Proponho situar, no ponto S barrado em relação ao pequeno a, o efeito fantasístico. Sua característica é ser

> uma relação articulada e sempre complexa, um roteiro que pode permanecer latente por muito tempo num certo ponto do inconsciente, mas que, não obstante, é organizado - assim como um sonho, por exemplo, só é concebido quando a função do significante lhe confere sua estrutura, sua consistência e, ao mesmo tempo, sua insistência (LACAN, 1999, p. 423).

Lacan traduz essa dinâmica desejante-fantasística com a fórmula $◊a, S barrado punção de a – a *lógica do fantasma* ou da *fantasia*. Punção aqui mobiliza uma rede significante que passa por furo, objeto pontiagudo que fura ou marca, barra de aço que grava letras numa tipografia, enquanto o símbolo ◊ transmite um certo efeito de mão-dupla.

> Temos aqui, em (S◊a), o correspondente e o suporte do desejo, o ponto em que ele se fixa em seu objeto, o qual, muito longe de ser natural, é sempre constituído por uma certa posição do sujeito em relação ao Outro. É com a ajuda dessa relação fantasística que o homem se encontra e situa seu desejo. Daí a importância das fantasias (LACAN, 1999, p. 455).

O psicanalista MD Magno ressalta os dois níveis da fantasia: há a "fantasia primordial", "Haver desejo de impossível", o nível absolutamente geral e genérico em que o movimento pulsional busca um gozo absoluto absolutamente impossível. Nesse nível o que há é a "*alucinação* de um objeto impossível, que não há" (MAGNO, 2010, p. 279). Mas essa fantasia primordial se inscreve em cada um como traço unário, como fixação do movimento desejante numa fantasia particularizada, singularizada. "A fantasia, $◊a, colocada por Lacan (...) só pode ser a tradução declinada, (...) para cada um, do que se inscreve como fantasia primordial" (MAGNO, 2010, p. 280). A partir dessa fixação singularitária vão se compor as tais cenas, roteiros, histórias que, para cada um, no nível particularizado da fantasia, prometem o gozo absoluto requisitado pela fantasia primordial.

Glynos e Howarth descrevem com clareza a dimensão sociopolítica da fantasia. Fantasias sociopolíticas assumem a possibilidade de identificação plena de um "todos nós" em torno das mesmas definições e objetivos. Essa identificação é vendida como possibilidade de realização da onipotência – o gozo absoluto desejado – ao mesmo tempo em que promete proteção contra o encontro traumático com o impossível – a impotência absoluta advinda do fato de que o gozo absoluto desejado é absolutamente impossível. Para tanto, a fantasia primordial – o gozo absoluto – é novamente particularizada, mas dessa vez em torno de algum tipo de significação sociopolítica, enquanto o trauma do impossível é velado pela particularização em torno de um inimigo diabólico apontado como causa da impossibilidade da satisfação absoluta.

> Seja no contexto das práticas sociais ou das práticas políticas, a fantasia opera de modo a ocultar ou "bloquear" a contingência radical das relações sociais. Isso é feito através de uma lógica ou narrativa fantasmática que promete uma plenitude-por-vir, uma vez que um obstáculo implícito ou nomeado é superado – a dimensão beatífica da fantasia –, ou que prevê o desastre se o obstáculo se revela insuperável, que pode ser denominado a dimensão terrífica da fantasia. Por exemplo, imagens de onipotência ou de controle total poderiam representar a dimensão beatífica, enquanto imagens de impotência ou vitimização poderiam representar a dimensão terrífica de tentativas fantasmáticas para alcançar ou manter o fechamento (GLYNOS; HOWARTH, 2018, p. 64).

Desconstrução e psicanálise põem em cena outra experiência da linguagem. Linguagem não é princípio de reunião, mas de separação, divisão e singularização. Rede não-linear de interações, a linguagem também não oferece nenhum "fio" que possa ser seguido até um fundamento absolutamente válido. Como diz Lacan, "não há Outro do Outro", isto é, "não há metalinguagem" (LACAN, 2003, p. 325). A suposição de fundamentação absoluta não passa de mais uma fantasia primordial de onipotência. Nessa experiência da linguagem atravessada pela fantasia não há também qualquer possibilidade de "reencontrar" ou estabelecer um sentido pré-determinado para "todos nós".

Segundo nos parece, o inquietante fenômeno contemporâneo da disseminação e adesão em massa às chamadas *fake news* é inseparável da fantasmaticidade sociopolítica enunciada por Glynos e Howarth. Notícias falsas que ganham adesão em massa apelam para fantasias de onipotência, de plenitude beatífica em torno de ideais e identidades comuns e alimentam a imagem horrífica de um outro diabólico. Assim, elas reforçam num nível grupal a fantasia primordial de gozo absoluto – e de proteção contra o trauma do impossível.

Segundo essas articulações teóricas, a adesão em massa a *fake news*, em geral lida como sintoma de uma "irracionalidade social" seria, pelo contrário, sintoma de uma *racionalidade social* extremada, que tenta finalmente reunir toda a sociedade em torno de um mesmo ideal de verdade e de uma mesma finalidade concebida como bem supremo. O gozo absoluto desse ideal supremo é impedido pelo próprio real – ou seja, é impedido por sua própria impossibilidade constitutiva. Por isso, para manter vivo o ideal, elege-se um "outro" demoníaco ao qual se atribui a verdadeira causa dessa impossibilidade. Esse outro deve ser firmemente combatido. Porém, note-se que, tanto melhor esse "outro" ocupa o lugar horrífico quanto mais difícil for sua erradicação. Assim, o combate pode/deve continuar indefinidamente, adiando eternamente o encontro com o real impossível.

Desconstrução e Psicanálise, ao investir numa compreensão não-logocêntrica da existência, isto é, a partir da diferença, da alteridade, da singularidade e da impossibilidade de reunião e pré-determinação plenas, exigem também um outro pensamento da Educação e do Currículo, que é o que veremos a seguir.

5. Um outro pensamento curricular

Vem se desenvolvendo nas últimas décadas um pensamento curricular inspirado pelas reviravoltas trazidas pela desconstrução e pela psicanálise. Procuraremos mapear nessa seção algumas das principais contribuições desse pensamento curricular (des)orientado pela singularidade, pela diferença, pela alteridade e pelo impossível.

Se as teorias curriculares do século XX tinham um forte sentido – logocêntrico – de reunião, de estabelecer um mesmo projeto comum a todos, para os teóricos desconstrucionistas não se trata nem de uma educação "comum para todos", mas nem também de uma educação "para cada um", personalizada ao gosto do cliente.

A defesa de um "*common core*" ou uma "base comum" que se materializa como política curricular na implementação da Base Nacional Comum Curricular entre 2017 e 2018 aposta na *reunião* de diferenças, alteridades e singularidades em torno de um "todos" homogêneo idealizado.

> Formular conteúdos curriculares comuns, a despeito do jogo político, pressupõe tratar todos os estudantes como iguais e, por isso, merecedores (ora necessitados, ora com direitos) dos mesmos saberes. Como fica, nessa perspectiva, a questão de que estamos sujeitos ao diferir que não pode ser homogeneizado? Estamos fadados à heterogeneidade que não permite afirmar, seja por merecimento, necessidade ou direito, quais saberes são/serão passíveis de estar conectados a essas múltiplas singularidades (LOPES, 2015, p. 457).

A "base comum" reforça a ideia de que o sentido dos processos educacionais é a transmissão de conteúdos de conhecimento entre sujeitos racionais e conscientes. Sejam conteúdos de um conhecimento compreendido como meramente intelectual, sejam conhecimentos compreendidos como dotados de um sentido prático, um "saber fazer", uma "competência". Essa centralidade de um certo ideal de "conhecimento" e da noção de "aprendizagem" como sentido primeiro

e último dos processos educacionais é o que Gert Biesta chama de "*learnification*", "aprenderismo": "a transformação do vocabulário utilizado para falar sobre educação em um vocabulário de 'aprendizagem'"[3] (BIESTA, 2016, p. 18). Seja como conteúdo meramente intelectual ou como aprendizado de um saber-fazer competente, o conhecimento é compreendido "como bem privado, algo passível de ser adquirido pelos sujeitos e utilizado para o alcance de determinados objetivos políticos, sociais, econômicos" (PEREIRA, 2017, p. 602), como objeto a ser "dominado" por um sujeito e então empregado com fins pré-estabelecidos. Como apontam Hugo Costa e Alice Casimiro Lopes:

> Supor um dado conhecimento como tendo onipotência funcional para todo contexto (da escola, do trabalho, da família, da sociedade, qualquer enunciação contextual) assinala uma tentativa de cálculo sobre as formas de conhecer o mundo, de lidar com o desconhecido, de aplacar o questionamento ignorado de uma alteridade "toda outra" que impõe, continuamente, a necessidade de revolvermos nossas formas de conhecimento, sejam elas quais forem (...). Não só o contexto não é algo calculável, como o conhecimento possível e passível de operação contextual não é uma propriedade carregada por um sujeito de razão/consciência transcendental, ou que possa ser causado por uma propriedade de conhecimento. (COSTA; LOPES, 2018, p. 317-318)

Essa padronização dos conteúdos a serem ensinados e das finalidades dos processos educacionais colocam-se também a serviço de dinâmicas avaliação externa que reforçam mecanismos de controle e regulação da docência:

> Currículo e processos de avaliação do desempenho têm estado presentes no Brasil e em outros países, em propostas e normativas orientadoras de políticas de currículo para a docência, como tentativas de controle da profissão. Operando com a organização curricular por competências e a avaliação do desempenho docente delineiam-se, em um contexto conservador, modos de significar o trabalho docente voltados à instrumentalização e padronização de percursos de formação inicial e continuada para fins de mensuração dos processos avaliativos. Tais processos engendram dispositivos de regulação desses profissionais que merecem nosso aprofundamento e problematização (DIAS; OLIVEIRA, 2021, p. 993).

Em *Ensinando por códigos: construindo uma docência padrozinada*, Rita Frangella coloca em questão a codificação presente na BNCC, das "habilidades e competências" a serem exercidas por professores e aprendidas por alunos, em "códigos alfanuméricos" que teriam o sentido de facilitar a transmissão e o aprendizado. No entanto, tais códigos aparecem como representantes por excelência de uma lógica do puro cálculo que é muito presente nas políticas curriculares. Há um saber prévio plenamente estabelecido. Basta ao professor consultar o código correspondente e aplicá-lo maquinicamente.

> Se num primeiro momento, os códigos podem parecer apenas uma questão organizacional, contudo, mais que isso, os códigos são usados como uma maquinaria para difundir e projetar a qualidade almejada, estabelecendo as regras para um cálculo maquínico que, na problematização feita, incide sobre a regulação do trabalho docente assentada numa normatividade impositiva. Os códigos dão uma visibilidade exaustiva à BNCC e costuram os nexos entre essa e políticas de avaliação, de formação docente, de material didático. (...) É a marca da presença, aqui indicada como originária, unidade total que significa um tempo homogêneo de repetição sucessiva. Busca-se reproduzir um discurso, contendo o efeito de incerteza que uma abertura a outras formas de pensar/fazer currículo poderiam fazer emergir, essas vistas como ameaça (FRANGELLA, 2021, p. 1165).

Frangella recorre a Derrida para pensar a educação como processo de abertura ao outro. Um outro que sempre escapa à calculabilidade:

> Imprevisível, indecidível, o outro que chega como acontecimento que excede o cálculo, transborda. É preciso rasurar os códigos, naquilo que ele tenta obliterar. A precisão e clareza que eles insinuam garantir são estratégias de fechamento, investimento numa unidade da totalidade. É essa unidade que precisa ser combatida, problematizada em favor da abertura, da ruptura, da singularidade (FRANGELLA, 2021, p. 1165).

[3] Tradução minha. Original em inglês: "learnification": "the transformation of the vocabulary used to talk about education into one of 'learning' and 'learners.'"

A padronização em torno de conteúdos comuns se coloca a serviço de projetos de "reconhecimento", como se o sentido ético-político de um processo educacional fosse tornar um sujeito propriamente reconhecível no interior da "sua própria" cultura (MACEDO, 2017, p. 547) – que se reconheça plenamente nela e que seja plenamente reconhecido por ela. Como afirma Elizabeth Macedo, "a experiência de reconhecimento, transformada em projeto pela teoria curricular, é uma violência ético-política" em relação às diferenças (MACEDO, 2017, p. 546). Pois assume-se uma homogeneidade cultural inexistente, bem como a suposição de que o sentido de um processo educacional e reforçar a "reunião" dos diferentes sujeitos nessa comum-unidade em torno de um padrão de identidade já estabelecido. Como nos lembra Elizabeth Macedo,

> talvez o exemplo mais significativo da tentativa de sancionar determinadas identidades via educação se faça com a ideia de cidadão. Ela se apresenta como o passaporte para toda e qualquer intervenção educativa, de perspectivas críticas — cidadão emancipado — a instrumentais em que cidadania é sinônimo de empregabilidade e capacidade de consumo (MACEDO, 2017, p. 546).

Nem mesmo o "reconhecimento" das diferenças é suficiente para dar conta da radicalidade da *différance*. Pois não se trata de reconhecer o outro enquanto outro segundo um padrão de identidade pré-estabelecido segundo o qual suponho saber quem é esse outro e como lidar com ele. Trata-se de repensar nossa forma de lidar com o outro "como se o conhecêssemos e soubéssemos plenamente sobre seu futuro, sua forma de pensar, experienciar o mundo" (COSTA, 2024, p. 22). Para além dos "reconhecimentos", tratar-se-ia então de cultivar uma abertura ao outro como "alteridade desconhecida, imponderável" (Costa, 2022, p. 04), nunca inteiramente apreensível, representável. Nossa lida com o outro é, segundo esse registro, sempre atravessada por um não-saber, por uma imprevisibilidade incontornável que se trataria então de *afirmar* e não de procurar eliminar.

Importante ressaltar que a oposição ao ideal logocêntrico do "comum" não se faz em defesa de uma educação personalizada ao gosto do cliente e sim de uma aposta na lida com o outro enquanto outro, assombrada pela indecidibilidade, uma aposta "na crença de que o público não significa a diluição do todos no um da nação (ou do mercado), mas o compromisso com deixar emergir a diferença concreta" (MACEDO, 2015, p. 905). Trata-se, portanto, de pensar "a educação desse sujeito que se constitui na relacionalidade com a alteridade, atravessado de múltiplas formas, nem sempre conscientes, por ela" (MACEDO; MILLER, 2022, p .13). Lidando com existências constituídas na – e desconstituídas pela – relacionalidade, os processos educacionais não podem ser personalizados porque não existe o indivíduo plenamente constituído com propriedades inerentes ou essenciais aos quais se teria de "atender". Nem mesmo se trata de "acolher", "reconhecer" ou simplesmente "respeitar" as diferenças. Isso ainda seria pouco. Trata-se de compreender que toda existência é efeito de relações com muitos outros diferentes que a atravessam de maneira nem sempre harmoniosa, nem sempre agradável, nem sempre pacífica. Uma educação (des) orientada pela diferença e pela singularidade é também uma educação que aposta na perturbação mútua entre diferentes e nas possibilidades impensadas que podem emergir dessas perturbações mútuas. Assim, "rasgar a ficção do sujeito proprietário e autoconsciente, é apenas para recuperar o potencial educativo do desfazimento do sujeito pelo outro que o atravessa, na escola e alhures, sem pedir licença" (MACEDO; MILLER, 2022, p. 14).

Para além da responsabilidade kantiana relativa ao "dever" de orientação pelo universal e para além da "responsabilização" e "prestação de contas" dos envolvidos em processos educacionais na forma da "*accountability*" (MACEDO, 2017, p. 511), os pensadores aqui elencados encarecem

uma noção derridiana de responsabilidade como resposta ao outro sempre singular "que nos interpela e que (...) não pode ser assimilado ou subsumido ao já dado" (MACEDO; SILVA, 2021, p. 56). Responsabilidade é a tentativa incalculável de fazer justiça à singularidade do outro – valendo lembrar, sempre que possível, que "nós mesmos" somos também "outros", singularidades incalculáveis e inapreensíveis por algum "centro" de identidade.

É em nome da abertura ao outro, à alteridade, à singularidade que Alice Casimiro Lopes coloca em questão o sentido fortemente racionalista tradicionalmente assumido pelas teorias curriculares: "A história do currículo é marcada pela ideia de que possa existir uma base racional que sustente as decisões sobre os saberes e atividades de ensino", seja ela "em função de princípios epistemológicas, psicológicos, ou mesmo emancipatórios" (LOPES, 2015, p. 455). Essa base racional é o fundamento logocêntrico que pré-determina o sentido e a finalidade dos processos educacionais. Partindo de inspirações desconstrucionistas, Lopes defende um "currículo sem fundamentos", isto é, um questionamento do logocentrismo curricular em nome de uma abertura às diferenças e singularidades que, em seu jogo conflituoso, negociam incessantemente os sentidos e significações envolvidas nos processos educacionais.

> Defender um currículo sem fundamentos remete à defesa de que não há princípios e regras absolutos, definidos cientificamente ou por qualquer outra dada razão, fora do jogo político educacional, que nos façam supor ser possível descansar da negociação de sentidos (LOPES, 2015, p. 462).

Em contraste com os projetos curriculares totalizantes e universalizantes, à maneira do racionalismo aristotélico e kantiano, essas teorias curriculares – que se apresentam como teorias curriculares pós-estruturalistas – tomam como ponto de partida a impossibilidade de reunião em torno de um mesmo "comum" a "todos". O "real" lacaniano (LOPES; BORGES, 2015, 2021; LOPES, 2021) e a *différance* derridiana (MACEDO, 2015, 2017, 2018; LOPES, 2018; MACEDO; RANNIERY, 2023; FRANGELLA, 2021) são tomados como marca dessa impossibilidade constitutiva de "reunião". Tais teorias colocam em questão o puro cálculo de um fundamento e uma finalidade pré-estabelecidos como absolutamente válidos em nome de uma abertura à alteridade e a um porvir incalculáveis. Não apresentam, no entanto, uma nova proposta curricular que pudesse substituir as anteriores, se estabelecendo como norma padrão aplicável "para todos" como política curricular vigente. O próprio pensamento que mobilizam permanece estranho a esse desejo de totalização. Trata-se de colocar em questão as padronizações curriculares de viés racionalista para abrir espaço para experimentações diversas, imprevisíveis e incalculáveis, na relação com os muitos "outros" sem os quais não há processos educacionais.

Considerações finais

Neste artigo tentamos ressituar a questão da racionalidade e da irracionalidade sociais a partir de alguns elementos da desconstrução e da teoria psicanalítica. Começamos por problematizar a própria noção de "racionalidade" como herdeira dos princípios logocêntricos de "reunião" e de orientação por princípios e finalidades pré-estabelecidas. Em seguida trouxemos a questão da adesão em massa a *fake news* para pensá-la para além do registro da "irracionalidade social" a partir da noção psicanalítica de fantasia. Por fim, apresentamos teorias educacionais contemporâneas inspiradas pela desconstrução e pela psicanálise procurando apontar suas diferenças em relação aos projetos educacionais logocêntricos.

Talvez já esteja suficientemente claro, mas talvez nunca seja demais ressaltar, que os projetos educacionais-curriculares que apresentamos, de inspiração desconstrucionista e psicanalítica,

não se fazem em nome de um "relativismo" irrefreável que inviabiliza qualquer projeto de "verdade", lançando a existência no caos, no "irracionalismo" e na mentira – cenário que nos tornaria presas fáceis, então, dos autoritarismos mais nefastos. Insistimos aqui que não se trata de defender nenhum projeto "irracional" de educação ou de vida, bem como insistimos em ressaltar que *relacionalismo não é sinônimo de relativismo*. Pelo contrário: o cultivo de uma sensibilidade relacionalista se faz em franca oposição a todo e qualquer projeto de fundamentação absoluta – os quais, por sua vez, tendem a ser projetos de fixação de identidades e de normas éticas no nível do cálculo. Uma sensibilidade relacionalista envolve também uma defesa intransigente das diferenças enquanto diferenças, das alteridades enquanto alteridades, das singularidades enquanto singularidades, isto é, enquanto inapreensíveis por padrões de identidade, verdade ou moralidade com pretensões de fixidez, universalidade e homogeneidade.

Se por um lado isso nos parece muito distante do "relativismo", arriscaria dizer que, por outro lado, uma sensibilidade relacionalista tende a nos colocar sempre a uma certa distância de projetos sociopolíticos autoritários. Projetos autoritários, afinal, tendem a se apresentar como defensores de fundamentos absolutamente verdadeiros, de parâmetros morais que se supõem absolutamente válidos e de modelos ideais de sociedade que devem ser recuperados de um passado idílico ou construídos num futuro utópico – bastando que se combata com o devido empenho os "outros" diabólicos que supostamente impedem a concretização de tais projetos.

Não compartilhamos a crença de que o "antídoto" para o fenômeno contemporâneo da adesão em massa aos autoritarismos e às *fake news* seja uma educação (re)orientada por ideais de verdade e racionalidade. Compartilhamos, pelo contrário, com Paulo Cesar Duque-Estrada, a percepção de que

> O problema, com toda a sua gravidade e relevância, é que o obstinado desejo de verdade não converge para reparar, mas sim prolongar e mesmo acirrar a *destrutividade em curso* – ontem assim como hoje – da chamada vida social. Destrutividade entendida aqui em dois sentidos muito precisos: por um lado, destruição de tudo o que lhe for diferente, estrangeiro, de outra ordem; por outro, destruição de si mesma, autodestruição. Para dizer de outro modo, a afirmação ou a consolidação de uma verdade única, estável e igualmente aplicável a tudo e a todos, constitui sempre (...) um golpe violento, uma injustiça, uma forma de anulação, abafamento, repressão, dirigida à diversidade, à heterogeneidade, às diferenças nas quais e através das quais tudo se ergue, se tece e acontece.(DUQUE-ESTRADA, 2020, p. 109).

Respondendo, portanto, ao protagonista do curta de Godard, talvez a "terrível mecanicidade" – que não creio, aliás, ser privilégio do nosso tempo – esteja muito menos associada a uma "morte da lógica" do que a um *excesso de lógica*, um excesso de racionalidade social ou, para usar o termo de Martins em seu estudo sobre as *fake news*, a uma "hipertrofia da verdade" (MARTINS, 2020).

Algo parecido com um "antídoto" contra projetos totalizantes, segundo cremos, não passa nem pela racionalidade nem pela irracionalidade. O antídoto contra projetos totalizantes em torno de sentidos de veracidade e moralidade supostamente absolutos talvez tenha mais afinidade com o exercício que o psicanalista MD Magno atribui à própria psicanálise. Segundo Magno "a psicanálise é o exercício da decepção" (MAGNO, 2018, p. 119). É por acolhermos a experiência da decepção diante de promessas de gozo absoluto que aprendemos a suspeitar quando elas assumem novas vestimentas, novas bandeiras e tentam se vender para nós novamente. Não é por esclarecimento ou moralidade que se nega tais projetos absolutizantes e totalizantes, é por guardar em algum lugar o resquício do gosto amargo da decepção.

No entanto, embora tenhamos evocado aqui espectros, fantasmas, sombras, impossibilidades,

decepções e um certo amargor, nada disso tem o intuito de promover paralisia ou uma desistência em relação a qualquer tentativa de criação de – ou insistência em – sentidos e experimentações que nos pareçam mais interessantes do que os socialmente disponíveis. Não é em nome da paralisia que falamos, mas justamente em defesa do movimento. Como bem dizem Borges e Lopes:

> Há quem tente se orientar frente ao abismo da falta de fundamentos buscando construir um fundo aparentemente sólido no qual se apoiar. Essa solidez, porém, é decorrente de sedimentos superpostos em um meio aquoso no qual submergimos tentando em vão buscar um ponto que nossos pés alcancem. Sugerimos que paremos de buscar o chão e comecemos a nadar, revolvendo esses sedimentos, turvando a água, mas ao mesmo tempo desestabilizando o que se apresenta estável e incontestável (BORGES; LOPES, 2015, p. 505).

É o que propomos com Borges e Lopes: uma afirmação da ausência de fundamentação absoluta como princípio capaz de (des)orientar posturas pessoais, políticas e educacionais em nome de uma abertura interminável à alteridade e à singularidade.

Referências

ARISTÓTELES. 2004. *Política*. Tradução de Therezinha Deutsch e Baby Abrão. São Paulo: Nova Cultural.

ARISTÓTELES. 2006. *A Política*. Tradução de Roberto Leal Ferreira. São Paulo: Martins Fontes.

ARISTÓTELES. 2002. *Metafísica*. Tradução de Marcelo Perine. São Paulo: Loyola.

BIESTA, G. 2016. *Good Education in the age of Measurement: ethics, politcs, democracy*. New York: Routledge.

BORGES, V.; LOPES, A. 2015. Formação docente, um projeto impossível. *Cadernos de Pesquisa*, São Paulo, v. 45 n. 157, p. 486–507, jul-set.

BORGES, V.; LOPES, A. 2021. Por que o afeto é importante para a política? Implicações teórico-estratégicas. *Práxis Educacional*, Vitória da Conquista, v. 17 n. 48, p. 1-22, out-dez.

CASSIN, B.; AUVRY-ASSAYAS, C.; ILDEFONSE, F.; LALLOT, J.; LAUGIER, S.; ROESCH, S. 2014. *Dictionary of untranslatables*. Tradução de Emily Apter, Jacques Lezra e Michael Wood. New Jersey: Princeton University Press.

COSTA, H. 2024. Como combater políticas nefastas? Currículos nacionais, formação de professores e educação em Geografia. *Revista Brasileira de Educação em Geografia*, Campinas, v. 14 n. 24, p. 05-26, jul-dez.

COSTA, H.; LOPES, A. 2018. A contextualização do conhecimento no ensino médio: tentativas de controle do outro. *Educação & Sociedade*, Campinas, v. 39 n. 143, p. 301-320, abr-jun.

COSTA, H.; LOPES, A. 2022. O conhecimento como resposta curricular. *Revista Brasileira de Educação*, Rio de Janeiro, v. 27 e270024, p. 1-23.

DERRIDA, J. 1991. *Margens da Filosofia*. Tradução de Joaquim Costa e António Magalhães. Campinas: Papirus.

DERRIDA, J. 2012. Uma certa possibilidade im-possível de dizer o acontecimento. Tradução de Piero Eyben. *Revista Cerrados*, Montes Claros, v. 21 n. 33, p. 229-251.

DERRIDA, J. 2010. *Força de Lei*. Tradução de Leyla Perrone-Moisés. São Paulo: Martins Fontes.

DERRIDA, J.; FERRARIS, M. 2001. *A Taste for the Secret*. Tradução de Giacomo Donis. Cambridge: Polity Press.

DERRIDA, J. 2018. "É preciso comer bem" ou "o cálculo do sujeito". Tradução de Carla Rodrigues e Denise Dardeau. *Revista Latinoamericana do Colégio Internacional de Filosofia*, Valparaíso, n. 3, p. 149-185.

DIAS, R.; OLIVEIRA, M. 2021. Dispositivos de regulação da docência nas políticas de currículo. *Revista Currículo sem Fronteiras*, Pelotas, v. 21 n. 3, p. 992-1000, set-dez.

DUQUE-ESTRADA, P.C. 2020. *Estudos ético-políticos sobre Derrida*. Rio de Janeiro: Mauad/PUC-Rio.

DUQUE-ESTRADA, P.C. 2004. Alteridade, Violência e Justiça: Trilhas da Desconstrução. In:

DUQUE-ESTRADA, P.C. (Org.). *Desconstrução e Ética*. Rio de Janeiro: PUC-Rio; São Paulo: Loyola.

FRANGELLA, R. 2021. Ensinando por códigos: construindo uma docência padronizada. *Revista Currículo sem Fronteiras*, Pelotas, v. 21 n. 3, p. 1148-1168, set-dez.

GLYNOS, J.; HOWARTH, D. 2018. Explicação crítica em Ciências Sociais: a abordagem das lógicas. In: LOPES, A.; OLIVEIRA, A.; OLIVEIRA, G. (orgs). *A Teoria do Discurso na Pesquisa em Educação*. Recife: Ed. UFPE.

GODARD, J.L. 1962. *Le nouveau monde*. Itália/França, 20 min.

HEIDEGGER, M. 2002. *Caminho de Floresta*. Tradução de Alexandre de Sá. Lisboa: Fundação Calouste Gulbenkian.

KANT, I. 2017. *A metafísica dos costumes*. Tradução de José Lamengo. Lisboa: Fundação Calouste Gulbenkian.

KANT, I. 2016. *Crítica da Razão Prática*. Tradução de Valério Rohden. São Paulo: Martins Fontes.

KANT, I. s/d. *Ideia de uma história universal com um propósito cosmopolita*. Tradução de Artur Morão. Disponível em: www.lusofia.net . Acessado em 06 jul. 2025.

KANT, I. 1999. *Sobre a Pedagogia*. Tradução de Francisco Fontanella. Piracicaba: UNIMEP.

LACAN, J. 1985. *O Seminário, livro 02: o Eu na teoria de Freud e na técnica da psicanálise*. Tradutores Marie Christine Lasnik Penote e Antonio Quinet. Rio de Janeiro: Jorge Zahar.

LACAN, J. 1995. *O seminário, livro 04: A relação de objeto*. Tradução de Dulce Duque Estrada. Rio de Janeiro: Jorge Zahar.

LACAN, J. 1999. *O seminário, livro 5: As formações do inconsciente*. Tradução de Vera Ribeiro. Rio de Janeiro: Zahar.

LACAN, J. 2008. *O Seminário, livro 14: A lógica do fantasma*. Tradução de Letícia Fonsêca. Recife: Centro de Estudos Freudianos do Recife.

LACAN, J. 2003. *Outros escritos*. Tradução de Vera Ribeiro. Rio de Janeiro: Jorge Zahar.

LOPES, A.; MACEDO, E. 2011. *Teorias de Currículo*. São Paulo: Cortez.

LOPES, A. 2015. Por um currículo sem fundamentos. *Linhas Críticas*, Brasília, v. 21 n. 45, p. 445–466, mai-ago.

LOPES, A. 2018. Sobre a decisão política em terreno indecidível. In: LOPES, A.; SISCAR, M. (org.). *Pensando a política com Derrida:* responsabilidade, tradução, porvir. São Paulo: Cortez.

LOPES, A. 2021. Radical investment in the curriculum in times of Covid-19: Can we question the anti-science discourses?. *Prospects*, Genebra, v. 51, p. 95–102.

MACEDO, E. 2017. Mas a escola não tem que ensinar?: Conhecimento, reconhecimento e alteridade na teoria do currículo. *Currículo sem Fronteiras*, Pelotas, v. 17 n. 3, p. 539-554, set-dez.

MACEDO, E. 2015. Base nacional comum para currículos: direitos de aprendizagem e desenvolvimento para quem?. *Educação e Sociedade*, Campinas, v. 36 n. 133, p. 891-908, out-dez.

MACEDO, E. 2018. A teoria do currículo e o futuro monstro. In: LOPES, A.; SISCAR, M. (orgs.). *Pensando a política com Derrida*: responsabilidade, tradução e porvir. São Paulo: Cortez.

MACEDO, E; MILLER, J. 2022. Por um currículo -Outro-: autonomia e relacionalidade. *Currículo Sem Fronteiras*, Pelotas, v. 22 e1153, p. 1-17.

MACEDO, E.; SILVA, P. 2021. Pesquisa pós-qualitativa e responsabilidade ética: notas de uma etnografia fantasmática. *Práxis Educacional*, Vitória da Conquista, v. 17 n. 48, p. 40–59, out-dez.

MACEDO, E.; RANNIERY, T. 2023. Derrida e a diferença: currículo como zona de tradução. *Imagens Da Educação*, Maringá, v. 13 n. 3, p. 26-46, jul-set.

MAGNO, MD. 2010. *Pedagogia Freudiana*. Rio de Janeiro: NovaMente Editora.

MAGNO, MD. 2018. *SoPapos 2016*. Rio de Janeiro: NovaMente Editora.

MARTINS, J.P. 2020. *A hipertrofia da verdade: Da vontade de verdade à vontade de identidade a partir de Nietzsche e Derrida*. Rio de Janeiro. 259p. Dissertação de Mestrado. Programa de Pós-Graduação em Filosofia da PUC-Rio.

NIETZSCHE, F. 2008. *A vontade de poder*. Tradução de Marcos Fernandes e Francisco de Moraes. Rio de Janeiro: Contraponto.

PEREIRA, T. 2017. Gramática e lógica: jogo de linguagem que favorece sentidos de conhecimento como coisa. *Currículo sem fronteiras*, Pelotas, v.3, p.600 – 616, set-dez.

SOLIS, D. 2014. Jacques Derrida e a frequentação dos espectros. In: HADDOCK-LOBO, R.; RODRIGUES, C.; SERRA, A.; AMITRANO, G.; RODRIGUES, F.; *Heranças de Derrida*: da ética à política. Rio d Janeiro: Nau.

As ciências de frente ao negacionismo, conspiracionismo e analfabetismo científico

The sciences in the face of negacionism, conspiracism and scientific illiteracy

Alexandre Luiz Polizel
Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo (IFES)
alexandre.polizel@ifes.edu.br

Abstract: As reflexões sobre as educações na contemporaneidade atravessam sintomas contemporâneos que representam disputas na configuração do campo simbólico-cultural e de suas linhas de subjetivação para a instauração dos modos de educar e formar sujeitos, dentre eles expressões reacionárias e neofundamentalistas. Dentre tais dinâmicas a disputa que tem sido realizada se dá no que toca o conceito e a significação das ciências. Desta percepção este ensaio objetiva-se em traçar as relações entre pensamento científico e teoria da cultura a partir de tais sintomas sociais contemporâneos, tomando o movimento Escola sem Partido como sintoma representativo. Tal escrita faz-se por uma hermenêutica Crítica e Clínica da cultura, compreendendo como linhas de constituição de tal sintoma o negacionismo, conspiracionismo e o analfabetismo científico.

Keywords: contemporaneidade; Crítica e Clínica da Cultura; educações; Escola sem Partido.

Resumo: Reflections on education in contemporary times cross contemporary symptoms that represent disputes in the configuration of the symbolic-cultural field and its lines of subjectivation for the establishment of ways of educating and forming subjects, including reactionary and neo-fundamentalist expressions. Among these dynamics, the dispute that has been taking place concerns the concept and meaning of science. From this perception, this essay aims to trace the relationships between scientific thought and cultural theory based on such contemporary social symptoms, taking the Escola sem Partido movement as a representative symptom. Such writing is done through a Critical and Clinical hermeneutics of culture, understanding denialism, conspiracism and scientific illiteracy as lines of constitution of such a symptom.

Palavras-chave: contemporaneity; Criticism and Clinic of Culture; education; School without a Party.

Recebido em 8 de julho de 2024. Aceito em 03 de novembro de 2025.

Introdução[1]

"Para dialogar é necessário pressupor uma gramática comum."(Wladimir Safatle)

Se pensamos nas possibilidades de uma vida em comunidade, em sociedade, mesmo em meio às premissas neoliberais de individualização dos sujeitos e de seus modos de ser, há de ser necessário imaginar um território que permita as relações. Tal território pode ser imaginado a partir do pensar um mundo em comum em seu campo de materialidades (MARX, 2013), de discursividades coletivas-coletivizadas (FOUCAULT, 2016), da corporificação do coletivo (DELEUZE; GUATTARI, 2011), do circuito dos afetos que interseccionam os corpos (SAFATLE, 2018) ou dos modos de subjetivação que os fazem coletivos (ROSE, 2011; FOUCAULT, 2010).

Tais aspectos requerem, segundo Vladimir Safatle (2017) e Jacques Derrida (1997), uma gramática comum, ou seja, uma possibilidade de elaborar modos de ler, escutar, enunciar e escrever o mundo de modo compartilhado[2]. Digo, só é possível imaginar um diálogo e o compartilhar o mundo, se uma gramática comum é possível. Há quem diga que tal aspecto soe com tom estruturalista, mas não nego a possibilidade de composições de estruturas – falhas, efêmeras, com fácil potencial do 'borrar' suas fronteiras – para enunciar mundos possíveis (AGAMBEN, 2019).

É deste perspecto de que há possibilidades de comuns que se funda a noção de escola. A escola como instância de representação dos espaços em que se compõe, compartilha e discute-se saberes, deriva-se da noção grega de *skholé*, que remete a uma "fonte de tempo livre" (MASSCHELEIN; SIMONS, 2013, p. 9), ou melhor, na ocupação de um tempo comum. Esta conceituação coloca em jogo o pensar que há um tempo e temáticas a serem ocupado s no espaço escolar, tendo assim seus fins formativos, para corpo-alma, dado coletivamente em uma gramática comum da ocupação do tempo.

O conceito de escola ao longo da história derivou, passou a ter outros traços e sentidos, como na Idade Média a ideia de uma ocupação de hermenêutica comum (AGAMBEN, 2019) e na modernidade da ocupação de currículos comuns (SILVA, 2015), mas tal composição sempre foi atravessada pela noção de uma gramática que poderia ser coletiva.

Não desejo, com isso, aqui, agenciar desejos a pensar em uma possibilidade de universalização, característica da operação de uma economia política dos corpos e mentes de matriz ideológica que subtraem a possibilidade de extensão e diferenciação para outros comuns (MASSCHLEIN; SIMONS, 2013). O que busco é ressaltar a indagação de que para dialogar, elaborar lações e se relacionar no circuito dos afetos (SAFATLE, 2018), é necessário imaginar que há gramáticas possíveis – e talvez impossíveis, mas que possam vir a ser (DELEUZE; GUATTARI, 1996; DELEUZE, 2018).

Todavia, a contemporaneidade é atravessada pela prefixação dos "pós-", são tempos de "pós-". Tal noção de "pós-" refere-se a um deslocamento da noção temporal, digo, o que é pós é, pois, veio a ser de um outro tempo e deixa seus sintomas no tempo presente em um vínculo com o

[1] Este artigo é parte constitutiva da investigação de doutoramento intitulada *Críticas e clínicas das culturas: educação, pensamento contemporâneo e o sintoma. Escola sem Partido*.

[2] Não gostaria neste momento de levantar a discussão se tal aspecto é possível, visto que as dinâmicas organizativas daquele que diz, das fantasias que atravessam a fala, o potencial de codificação e decodificação discursivo e as possibilidades de um fazer falar simétrico requerem outra discussão. Para tal, recomendo a obra de Vladimir Safatle, *Cinismo e a Falência da Crítica* (2008b).

passado, bem como o que é pós remete a uma promessa-profecia do que pode vir a ser, ou do que está vindo a ser (RICOEUR, 2010). Faz-se importante ressaltar que toda referência de tempo é também uma referência de espacialidade-territorialidade, logo, o deslocamento das possibilidades de compartilhar o tempo por estarmos sempre no mover-se dos "pós-" é também o arraste para a constante desterritorialização (DELEUZE; GUATTARI, 2011) e da quebra dos quadros referenciais (LATOUR, 1994; LIPOVETSKY, 2005).

Este abalo sísmico que conclama outra cronologia e topologia (LYOTARD, 2002) não se dá de forma aleatória, mas nos jogos de encontros e de forças que os produzem (NIETZSCHE, 2012; 1974). São diversos os jogos de força que localizam o cenário no campo dos "pós-", não por menos, ouvimos constantemente falar em pós-estruturalismo, pós-modernismo, pós-colonialismo… Nos jogos dos "pós-" o que se define é um espaço do não definível, ou da tentativa de indefinir algo, com intuitos discursivos da instauração de um cenário de crises: por compreender que neste algo novo se cria ou se permite destruir (COMITÊ INVISÍVEL, 2016). O tempo dos "pós-" é assim um tempo de cataclismas, e tomo o cataclisma como a expressão de rupturas e cisões.

Toda ruptura requer um exercício de força negativo, que nega, separa, suprime e afasta (NIETZSCHE, 2016; 2012; FREUD, 2014; 2010a; DELEUZE, 2018), e uma corporificação a qual se interessa em cindir. Percepto que a questão que permitiu o conectivo referencial-temporal-espacial da antiguidade à contemporaneidade inspirado por Michel Foucault (2016), é a questão da verdade. Tal aspecto é reiterado por pensadores como Bruno Latour (1994) e Immanuel Kant (2018), no que remete o que podemos conhecer, elaborar e ter um quadro referencial que passe por modos de veridicção – seja uma verdade de aspecto discursivo, factual e/ou categórico.

Assim, um dos elementos da contemporaneidade que evoca em seu circuito dos afetos os tempos de "pós-" é a desterritorialização e a descronologização invocadas no cenário da pós-verdade. Há algo posto em questão, no que toca a verdade, que vem-a-ser de um passado e que devem constantemente em um futuro próximo no instante. Há algo no campo da verdade que é posto como narrativa para disputar o que é passível de ser pensado, ocupado ou restringido. As forças negativas de cisão são inclinadas, então, para o espectro do que estaria no plano da verdade, ou melhor, no que seria jogado ao amálgama da pós-verdade.

O referido aspecto pode ser considerado ingênuo ou inocente, todavia tem-se por perspectiva que a apreensão pela ocupação de um tempo comum (MASSCHELEIN; SIMONS, 2013), uma hermenêutica coletiva (AGAMBEN, 2019) e/ou um currículo (SILVA, 2015), como um caminho possível, se deu pela ideia de que estas possibilitariam o alcance das verdades – ou dos saberes verdadeiros-validados por um critério de verdade – as possibilitariam o alcance de uma gramática comum para diálogos possíveis (SAFATLE, 2017).

Todavia, "o traço maior da subjetividade em tempos de pós-verdade será exatamente esta aptidão para a inversão" (DUNKER, 2017, p. 13) ao passo que se as possibilidades de verdades se mostravam uma gramática comum, agora elas representam seu inverso, a ausência das gramáticas. Christian Dunker (2017) nos ajuda a refletir tal aspecto ao perceber que as 'questões das verdades' têm em sua base três conotações comuns: "a revelação grega (*alethéia*) de uma lembrança esquecida quanto à precisão latina do testemunho (*veritas*) e ainda a confiança-judaico cristã da promessa (*emunah)*. Por isso a verdade tem três opostos diferentes: a ilusão, a falsidade e a mentira" (DUNKER, 2017, p. 18). Contudo, nos tempos de "pós-" estas raízes rizomáticas-arbóreas

se invertem e desnaturam, ao passo que três técnicas-operações são acionadas: negacionismo, conspiracionismo e o analfabetismo científico.

Em minhas buscas de escuta do que me diz o sintoma Escola sem Partido, identifica-se em seus efeitos operatórios estes três mecanismos. Assim, o intuito deste eixo analítico é apresentar as diferenciações destes elementos em termos filosóficos, pedagógicos e seus impactos nas educações e no pensamento científico contemporâneo. Neste, pretendo traçar as relações entre pensamento científico e teoria da cultura a partir de tais sintomas sociais contemporâneos.

Nego: negação, negacionismo e perversão

Nas buscas pela escuta de algo que nos diz o sintoma Escola sem Partido, nos deparamos com sua estruturação a partir da negação. O projeto de Lei (BRASIL 2019; 2016; 2015) que busca estabelecê-lo enquanto mote para regulamentar restritivamente as dinâmicas educacionais de uma ocupação do tempo-temas-reflexões comuns, se baseia na lógica proibitiva do negar, vide:

> I – *não* se aproveitará da audiência cativa dos alunos para promover os seus próprios interesses, opiniões, concepções ou preferências ideológicas, religiosas, morais, políticas e partidárias;
> II – *não* favorecerá nem prejudicará ou constrangerá os alunos em razão de suas convicções políticas, ideológicas, morais ou religiosas, ou da falta delas;
> III – *não* fará propaganda político-partidária em sala de aula nem incitará seus alunos a participar de manifestações, atos públicos e passeatas;
> IV – ao tratar de questões políticas, socioculturais e econômicas, apresentará aos alunos, de forma justa, as principais versões, teorias, opiniões e perspectivas concorrentes a respeito da matéria;
> V – respeitará o direito dos pais dos alunos a que seus filhos recebam a educação religiosa e moral que esteja de acordo com as suas próprias convicções;
> VI – *não* permitirá que os direitos assegurados nos itens anteriores sejam violados pela ação de estudantes ou de terceiros, dentro da sala de aula (BRASIL, 2019, p. 3, grifos meus)

A própria escala restritiva que situa o '*dizer não*' encontra-se baseada nos princípios da noção de moralidade individual (BRASIL, 2019) que, como toda estruturação moral, baseia-se em uma escala negativa (NIETZSCHE, 2019). Assim, o sintoma Escola sem Partido apresenta em sua estrutura de disposição supostamente legal-subjetiva uma ação dupla de negação: o dizer não ao Outro-educador a partir do dizer não de um Outro-moralizante. Não há, neste sentido, um critério de verdade que sustenta a tentativa de sedimentar-se como estrutura legal, visto que a estrutura legal deve ser traçada de forma positiva (FOUCAULT, 2008; LATOUR, 2013), ou seja, ela precisa apresentar um discurso afirmativo sobre aquilo que se diz e regulamenta o processo formativo. É no exato dizer e apresentar algo que se coloque enquanto termos e parâmetros evidenciáveis e que resistem ao julgo da crítica (FOUCAULT, 1990; SAGAN, 2006), que a concepção de verdade se coloca enquanto testemunha contrária a uma suposta ilusão.

Não quero, com isto, dizer que as verdades que se encontram em itens evidenciáveis devem ser preservadas como santos ocos em seus altares (NIETZSCHE, 1974), mas que a própria proposição de uma interpretativa-analítica da realidade requer como força produtiva um dizer afirmativo, um sim sobre o algo, e nesta há a possibilidade de "cria[r] o conceito de coisa" (NIETZSCHE, 2017b, p. 22). Em tal aspecto, a estrutura de uma dupla negação inverte duplamente a possibilidade da elaboração de uma gramática comum possível e, com isso, um primeiro deslocamento "pós-" em relação à verdade.

Para além disso, o dizer "não" remete a ao menos outros três aspectos no que refere a negação: a um negar de aspecto repressivo[3], ao negar de aspecto socio-subjetivo-ambiental[4] e ao negar de um aspecto perverso social.

A primeira dinâmica de negacionismo, seja em termos de frequência ou de instância de aparecimento, é aquela que se refere ao negar de um aspecto repressivo. Neste sentido o negacionismo teria sua primeira dinâmica de taxonomização-nomeação a partir da psicanálise, de modo que Sigmund Freud (2014, p.10-11) percebe que o processo de "negação é tomar conhecimento do reprimido; na verdade já e um levantamento da repressão, mas naturalmente a não aceitação do reprimido". De um olhar Freudiano percebe-se que o ato de negar, do dizer não, resvala na percepção de que há algo que se sabe no plano do recalcado, mas em um ato de mantê-lo negado diz não. Em outras palavras, é o modo de manter a percepção das coisas em um plano não consciente, ou de modo consciente diretamente recalcado.

São múltiplos os aspectos que levam ao processo de negação, sendo que três afetos são intercalares à operacionalização do negar: medo, ódio e a perda. Vemos neste horizonte que o primeiro elemento constitutivo do acionamento do processo de negação é o medo, visto que o medo indica um potencial de não referenciação de uma possível situação de amparo. O medo é o afeto constitutivo que coloca o real sob a óptica da dúvida, da crise, da desterritorialização e descronologização, do desamparo. O medo indica que algo pode modificar os parâmetros e transformar, desvelar ou criar outro cenário desconhecido. Frente à possibilidade de angústia e desamparo da perda de referências, o medo aciona o dizer não para os elementos que invocam esta realidade tortuosa. É nesta esteira que Vladimir Safatle (2018, p. 43) aponta-nos que com o "advento de uma sociedade da insegurança total, não muito distante daquela que podemos encontrar nas sociedades neoliberais contemporâneas [...] a circulação do medo [torna-se] afeto instaurador e conservador das relações", principalmente daqueles que buscam uma possível autoridade que os mantenha salvo e seguro do que lhes dá medo.

O medo, assim, torna-se elemento constitutivo de formulação de vínculos sociais daqueles que temem, bem como instala arcabouço para que um circuito dos afetos se funda e se associe ao desejo de autoridade (FREUD, 2017), de máquinas administrativas de normativas que afastem a possível fonte de contaminação (SAFATLE, 2018) e da perpetuação de dependência às respectivas normas a partir da noção de dever (CALLIGARIS, 2022).

Neste cenário, a autoridade vigente e as normativas passam a mobilizar um imperativo dos deveres para manutenção do cenário que rui – e que rui pela própria condição dos tempos de "pós-" –, sendo que se apenas o medo não se torna força mantenedora da topologia movediça do contemporâneo, outra linha de força negativa é acionada, o ódio. O ódio é afeto que mobiliza os corpos em direção à expressão da violência, visto que este permite lançar sobre o outro sua negação, deslegitimação, exclusão ou eliminação (AGAMBEN, 2004).

[3] Não estou aqui, neste texto, fazendo menção a uma hipótese repressiva, no sentido de recorrer a uma noção de relações de poder-desejo que não produzem algo, pelo contrário, corroboro com Vladimir Safatle (2008a) e Michel Foucault (2015), que as dinâmicas da economia libidinal são produtivas, até mesmo em seus fluxos de recalcamento e percepção consciente.

[4] Compreende-se que os fluxos de negação não são apenas em referência a uma temática, mas aos modos de subjetivação e de socialização que aquelas produzem, ou seja, as possibilidades e os modos de existência que podem ser instaurados a partir do que é negado (REICH, 1988). Esta percepção aqui tratada enquanto social-subjetiva-ambiental, é perceptada por Karl Manheim (1986) enquanto característica do pensamento conservador, daquilo que o mesmo percebe e nega com intuito de preservar um modo de vida vigente, ou para ressentir-se em relação aos modos de vida insurgentes (COMITE INVISÍVEL, 2016).

O ódio neste sentido torna-se a salva guarda dos mecanismos de proteção frente ao medo – autoridade-normativas-deveres –, pois garante licenciosidade para agir em relação a sua proteção (ou melhor, da proteção das referências em ruínas). A problemática que emerge é que o ódio é oposto à escuta, pois este renuncia a uma gramática comum e trata a gramática do Outro como desviada, sobra e possivelmente contaminante (TSING, 2019). No circuito dos afetos do medo e do ódio, não há espaço para diálogo, pois não se prescinde de gramática comum com uma possível fonte de todo mal.

É isto que vemos a exemplo nos textos do sítio eletrônico do Escola sem partido, tomando como representativos: *Caos: a receita de Jean Wyllys para a educação brasileira*[5]; *O pesadelo de Paulo Freire*[6]; *Totalitarismo através da Educação*[7]; A ideologia de gênero no banco dos réus[8]; *Alienação parental: uma prática cometida também pelos professores*[9]; entre outros.

Estes textos disponibilizados no sítio eletrônico nos fazem perceber que tal gramática comum, supracitada, não é prescindível pois instala um campo potencial de perda em seu processo de diferenciação (DELEUZE; GATTARI, 2011), ao passo que a diferenciação tem por preço sinalizar perdas (BIRMAN, 2019) e inaugura um cenário de mudanças-criações (NIETZSCHE, 2016). Há na intersecção entre o medo e o ódio um agente cimentante que como uma Moira sinaliza uma possível perda.

Há a problemática da perda um aspecto dual, de um lado a perda apresenta que aspectos de mudança acontecimental produziram uma ruptura de modo tal em que nada mais poderá ser como era antes, e de outro aspecto inaugura o cenário de que outro horizonte se cria e se produz (LATOUR, 2020). A perda, assim, sinaliza que há algo que se tornou impossível de repetir e, de outro, abre uma clareira onde algo impossível agora pode fazer-se possível de construção (MISKOLCI, 2018). Aquele que nega aspectos do real, pois não consegue lidar com o presente, mantém-se no negar de aspecto repressivo (por medo, ódio ou não percepção da perda), aqueles que tomam um novo campo do possível recaem na percepção de que as mudanças atravessam os aspectos subjetivos, sociais e ambientais (GUATTARI, 2009).

Esta abertura abre espaço para um segundo mote de negação, o negar de aspecto socio-subjetivo-ambiental. Este tipo de negacionismo remete a um aspecto que atravessa a ecologia dos modos de existência, ao passo de que há a percepção de que toda modificação atravessa as possibilidades de fazer-se sujeito (modos de subjetivação), as conformações sociais e os modos de existência possíveis, bem como os modos de compor com o ambiente-espaço. Estes três elementos são constitutivos dos territórios existenciais, ou seja, das formas e modos pelos quais se pode existir. A percepção de que os novos agentes emergem do espectro da diferença, se dá pela percepção de que o "pós-" anos 2000 "convidam a pensar uma encruzilhada, ou melhor, uma intersecção entre as diferentes formas de minoração do outro e de si mesmo, bem como as políticas de reversão dessa minoridade" (DUNKER, 2017, p. 16).

[5] Disponível em: <escolasempartido.org/blog/caos-a-receita-de-jean-wyllys-para-a-educacao-brasileira/>. Acesso em 20 de janeiro de 2023

[6] Disponível em: <escolasempartido.org/blog/o-pesadelo-de-paulo-freire/>. Acesso em 18 de janeiro de 2023

[7] Disponível em: <escolasempartido.org/blog/totalitarismo-atraves-da-educacao/>. Acesso em 20 de janeiro de 2023

[8] Disponível em: <escolasempartido.org/blog/a-ideologia-de-genero-no-banco-dos-reus/>. Acesso em 19 de janeiro de 2023

[9] Disponível em: <escolasempartido.org/blog/alienacao-parental-uma-pratica-cometida-tambem-por-professores/>. Acesso em 20 de janeiro de 2023

Digo, com isso, que as mudanças nas esferas éticas, estéticas, políticas, ontológicas, epistemológicas e tecnológicas que se dão na contemporaneidade reverberam modos outros de subjetivação, existência e territorialidade (MISKOLCI, 2018) de modo que a diferença se prolifera. Quando os sujeitos não conseguem negar em termos de aspecto repressivo, e não suportam as modificações subjetivas-sociais-ambientais, aciona-se um negacionismo de segunda via: nega-se os aspectos subjetivos-sociais-ambientais. Tal negação escapa o deslocamento da esfera da ilusão (contrária a *alethéia*) que fica em um plano de aspecto repressivo, para uma negação da ideia de falsidade (contrária a *veritas*), ao passo que se nega o próprio testemunho do real-acontecimental, tomando como falso até mesmo o que se encontra em um plano de evidenciação.

Tal aspecto, segundo Susan Faludi (2001), evoca diferentes formas de negacionismos atreladas aos *backslashes* e seus modos operatórios: estereotipação da diferença, inversão de causas e demandas desejantes responsabilizando a minoridade por problemas sociais sobre supostas perdas de referências, atribuição de características negativas às minoridades-diferenças, negação das experiências vividas que fazem pulular os devires, estratégias evocadas para garantia de uma negação dos movimentos-acontecimentos que elaboram discursividades e produzem modos de criação de outras subjetividades-sociedades-ambientes.

Há, nesta operação, a tentativa de negar os devires negando a diferença-minoridade. Esta negação se dá em ao menos dois movimentos: a tentativa de operacionalizar relações de poder diretivas para frear a proliferação das diferenças, agenciando os aspectos repressivos de medo e ódio e reverberando práticas de violências; bem como a tentativa de um segundo aspecto negativo-negacionista, que envolve uma postura não erótica, a negação enquanto aspecto perverso social.

Esta percepção me é colocada em interlocução com os escritos de Contardo Calligaris (2022, p. 43), ao passo que este percebe em seus escritos sobre o *Grupo e o Mal*, que a perversão sexual se encontra desarticulada da perversão social, ou seja, no caso da perversão social é evidente que opera "uma dinâmica assassina tanto ou mais violenta em seus efeitos, mas na qual, curiosamente, a grande massa dos assassinos não parece se motivar por qualquer forma de ódio". Tal aspecto se dá por múltiplos fatores: a não compreensão do que é de fato o risco posto, a não percepção do que se tratam as temáticas-enunciados-discursos, a não compreensão do que se está sendo discutido, a desorganização subjetiva frente ao tempo vigente, ao sentimento de poder frente a um coletivo que professa uma suposta ordem. A negação enquanto perversão social encontra-se, neste sentido, entranhada no imaginário coletivo, mas de forma desordenada-desregulada, de modo que o terceiro pilar que sustenta a possibilidade de verdade, a promessa (*emunah*), rui em não garantir oposição ao seu contrário (a *mentira*), ao passo que não se pode prometer nada na desordem se não o retorno à certeza do passado, e não se pode garantir que haverá resistência à potencial operação de mentira, se não há horizonte futuro para tal.

No que remete à promessa como o indicador de verdade que se mantém no tempo porvir, o ruir a possibilidade de promessa instala linhas de subjetivação que não se ancoram no futuro, mas prometem uma possível reestruturação no passado: moraliza-se o tempo que foi, deseja-o, mesmo sabendo que este se perdeu. Este retomar o passado não sustenta a oposição entre verdade e mentira, visto que não se faz possível a manutenção do regime de testes (SAGAN, 2006) em relação a um tempo que não pode ser acessado se não na memória e registros – mas em descrédito devido o aspecto repressivo que coloca em questão o testemunho –, bem como no qual não é possível colocar a verdade nas forças de torção que são necessárias para sustentação

do que estaria no plano da mentira (DELEUZE; GUATTARI, 1996; 1995; DELEUZE, 1988; FOUCAULT, 2016).

Vemos assim que a questão da "pós-" verdade é balizada pelo sintoma Escola sem Partido, ou ao menos expressa, desestabilizando os pilares do que permitiria pensar uma gramática comum em relação aos regimes de verdades possíveis (FOUCAULT, 2016) e quadros de referências que poderiam ser instalados (LATOUR, 2011), mas a desestabilização dos parâmetros de verdades pelos processos negativos do negar de aspectos repressivos (coloca em questão a *aletheia* e a ilusão), negar aspectos subjetivos-sociais-ambientais (coloca em questão a *veritas* e a falsidade), e da negação enquanto aspectos perverso social (coloca em questão a *emunah* e a mentira), vemos que na operação negativa-negacionista, no dizer não, o campo da gramática comum rui.

Conspiro: conspiracionismo, paranoia e o ressentimento no forjar do Outro

"Sempre há de encontrar um culpado conveniente para inocentar um herói ressentido." (Maria Rita Kehl)

Aponto aqui que há no sintoma Escola sem Partido algo que nos dá indícios da deterioração de uma gramática comum e, se o faço, situo a negação como uma das bases que esfacelam e desorganizam esta gramática coletiva ao deslocar a possibilidade de pensar em questão de 'verdades', situando-nos em um cenário da pós-verdade. Arrisco-me a chamar, no eixo anterior, que este processo operaria pela técnica da negação e dos negacionismos. Todavia esta não é a única instrumentalização – e aqui faço um jogo entre a razão instrumental (ADORNO; HORKHEIMER, 1991) e a paixão instrumental (CALLIGARIS, 2022) – e técnica de poder que é acionada-expressa neste sintoma. Situaria uma segunda técnica-tecnologia lançada pelo sintoma Escola sem Partido, com pontos de consonância ao negacionismo, mas que se diferencia em seus moldes operatórios e de subjetivações: o ato de conspirar.

A conspiração, o ato do conspiro, tem como elemento comum de primeira instância (em consonância ao negacionismo) ser um ato de negação. Vemos isso nos escritos de Sigmund Freud (1988), em suas *Observações psicanalíticas sobre um caso de paranoia* (*dementia paranodes*), o ato de conspirar encontra-se interligado a uma estrutura paranoide que nega algo que se encontra no plano do incompreensível e/ou não simbolizável. Contudo, tal aspecto diferencia-se do efeito apenas da negação, pois há investimento psíquico de buscar compreender-simbolizar aquilo que nega. Digo, com isto, que o 'conspiro' emerge de uma negação, mas se coloca a falar sobre aquilo que lhe é incompreensível para buscar compreendê-lo (DUNKER, 2003).

As tentativas de elaborar tal compreensão incorrem ao olhar para múltiplas identificações do que poderia indicar o incognoscível: a indicação de um processo persecutório por um Outro, a ideia de um adoecimento comum ou coletivo provocado por um Outro, a formulação de uma possível perda de potência causada por um Outro, a afirmação em uma posição de grandeza para enfrentar um Outro (DUNKER, 2003), a acusação de que um Outro seduz e perverte (KEHL, 2020), a massiva produção de neologismos e de ideias delirantes. Acionamentos que têm em seu funcionamento um ponto em comum, diz de um Outro, que diz de algo que teoricamente encontra-se fora do Eu.

Vemos que há a invocação deste Outro que espreita, planeja e possivelmente corrompe nos textos: O *Jornalismo a serviço da doutrinação*[10]; *O governo que nos educa*[11]; *Marxismo: ideologia oficial da escola pública de Santa Catarina*[12]; *O objetivo é doutrinar*[13]; *Plano Nacional de Educação irá aprofundar doutrinação no ensino*[14]; entre outros.

Este Outro-fora-do-Eu, torna-se ponto de ancoragem como intermediário (ou fim) para identificar a fonte do incompreensível-não-simbolizável. Este Outro pode expressar-se por múltiplas vias, sendo um outro materializável ou imaterializável, individual ou agrupamento, institucional ou difuso, perene ou efêmero, ideal ou ideológico, discursivo ou não discursivo (MACGOEY, 2019). Todavia, este Outro é instaurado na existência ao ser enunciado enquanto Outro (FOUCAULT, 2016), sendo que a força conspiratória situa-se em um duplo efeito agonístico de nomear este Outro: uma primeira instância negativa de enunciar que este outro 'não' deveria estar aqui, e por isso deveria ser combatido (o que lhe corporifica); e uma segunda instância afirmativa ao narrar que este Outro que se encontra aqui (afirma-o), não deveria estar (nega-o).

As intencionalidades que demarcam o conspirar contra o Outro que 'não deveria estar aqui', para os conspiradores, investem em uma lógica de natureza paranoide que se faz de dois tipos (não necessariamente desarticulados): paranoia de reivindicação e/ou paranoia de autopunição (LACAN, 1987; 1985). Tais categorias não são elementos fixos, mas se configuram enquanto estádios, processamentos e operações da subjetividade e dos motes de significação naquilo que toca modos de ser-estar-agir-pensar (ALLOUCH, 1997).

Esta operação do funcionamento subjetivo-paranoide resvala não em uma perspectiva individual (FREUD, 2017), mas é de instância e funcionamento coletivo (CANETTI, 2019), visto que os modos de subjetivação e operação em planos micropolíticos e macropolíticos tornam-os coletivizados e coletivizantes (FOUCAULT, 2016), bem como massificam a instrumentalização do conspiracionismo. Neste cenário da percepção paranoica "O sujeito encontra-se persuadido de que é vítima de um prejuízo e que sua causa é um complô dirigido contra ele. Sem outros fenômenos interpretativos, essa convicção ordena a sua construção delirante. Ultrapassando o interesse que toma na realidade, esse prejuízo ameaça sua própria existência enquanto *parlêtre*" (HAMON, 2020, p. 299).

Assim, é na percepção como (possível) vítima do Outro que o sujeito que conspira coloca-se na posição paranoide de reivindicar o direito à defesa, seja esta por meios institucionais, discursivos ou violentos. O direito à defesa operaria enquanto uma reivindicação subjetiva-desejante de eliminar o Outro, antes que este Outro atente contra si, mesmo que este outro não atente no plano real, o sujeito paranoide-conspirador faz a reivindicação em nome de um Ideal (HAMON, 2020). Esta idealidade lhe conferiria, inclusive, o direito do uso da violência contra o Outro (CANETTI, 2019). Esta violência não seria necessariamente de um cunho erótico – em que o sujeito goza ao reivindicar com a violência –, mas estaria mais interligada a uma paixão instrumental que opera

[10] Disponível em: <escolasempartido.org/blog/o-jornalismo-a-servico-da/>. Acesso em 2 de fevereiro de 2023

[11] Disponível em: <escolasempartido.org/blog/o-governo-que-nos-educa/>. Acesso em 2 de fevereiro de 2023

[12] Disponível em: <escolasempartido.org/blog/marxismo-ideologia-oficial-da-escola-publica-de-santa-catarina/>. Acesso em 2 de fevereiro de 2023

[13] Disponível em: <escolasempartido.org/blog/o-objetivo-e-doutrinar/>. Acesso em 2 de fevereiro de 2023

[14] Disponível em: <escolasempartido.org/blog/plano-nacional-de-educacao-ira-aprofundar-doutrinacao-no-ensino/>. Acesso em 2 de fevereiro de 2023

por meio de uma perversão social (e aqui vemos mais uma vez a relação do conspiracionismo com o negacionismo).

Digo, com isso, que as dinâmicas desorganizativas que situam o sujeito na posição dos "pós-", do não compreensível e não simbolizável, fazem aqui uma operação não de cunho negativo apenas, mas de um aspecto reativo (NIETZSCHE, 1974; DELEUZE, 2018) frente ao Outro e suas elaborações discursivas e de mundo. Não apenas nega, mas reivindica o direito de reação violenta e persecutória frente ao Outro, sob a óptica de que há nesta reivindicação um direito legítimo de possível defesa.

Aqueles que se encontram articulados à massificação conspiratória, que opera pela lógica paranoide-conspiradora, se não reivindicam a diferença por sua posição subjetiva, a fazem por sua posição coletiva, visto que é acionado uma segunda via de paranoia: a paranoia de autopunição (LACAN, 1987; LACAN, 1985).

A paranoia de autopunição começa a ser organizada em termos de pensamento e discursividade a partir dos escritos de Sigmund Freud sobre *O Eu e o isso*, ao passo que se vê nas dinâmicas subjetivas da autopunição uma vinculação com a culpa e com a moralidade, visto que:

> A tensão entre as exigências da consciência e os sentimentos concretos do eu é experimentada como sentimento de culpa. Os sentimentos sociais repousam em identificações com outras pessoas, na base de possuírem o mesmo ideal do eu. A religião, a moralidade e senso social foram originalmente uma só e mesma coisa. [...] Mesmo hoje os sentimentos sociais surgem no indivíduo como uma superestrutura construída sobre impulsos de rivalidade ciumenta contra seus irmãos e irmãs. Visto que a hostilidade não pode ser satisfeita, desenvolveu-se uma identificação com o rival anterior (FREUD, 1969, p. 52).

Em tais parâmetros, a culpa e a moralidade enquanto demandas (das perversões) sociais remetem ao sujeito que não reivindica à reatividade a posição de identificação e/ou complacência com a possível perturbação da ordem pelo Outro. Se este encontra-se e mantém-se na posição subjetiva de identificação com o grupo que lhe demanda conspiração-reivindicação, e este não o faz, o investimento é tomado em formato de culpa e de autopunição.

Esta autopunição é dada em ao menos três sentidos: i) a percepção de uma imposição moral, de modo que se não pode-deseja-quer agir por reivindicação-reatividade o sujeito coloca-se a impor a si e aos outros um quadro de referência moral para manter-se nas forças negativas (NIETZSCHE, 2019); ii) a idealização de um Eu, supostamente moralizado que se culpa pelo cenário vigente e, pelo ato de se culpar, tem por força reativa de seu ideal de Eu reagir culpando o que o Outro fez com o mundo, com os outros e consigo, o porquê permitiu que isto ocorre-se (FREUD, 1969); e iii) na percepção de que o ato de conspirar revela a necessidade de um agir intenso e incessante – e aí é morada de sua autopunição – que o requer posicionar enquanto bastião da moral e pela salvação em nome do coletivo, contra este Outro que se quer é de fato simbolizável e cognoscível (FERNANDES, 2001), mas que deve ser combatido com todas as forças aos limites do próprio esgotamento (HAN, 2018).

A conspiração-paranoide de autopunição, mesmo sem reivindicar, dependeria em seu autopunir um norte para direcionamentos das forças punitivas: o Outro. Este ato manter-se-ia no campo da conspiração pois há um Outro, externo a mim, que me faz me autopunir. Esta percepção da autopunição, pode, assim, desencadear em uma dinâmica de reivindicação e/ou em uma instrumentalização da paixão de modo reivindicativo-autopunitivo. Esta paixão-afeto torna-se instrumental pois opera enquanto um instrumento que movimenta as engrenagens, técnicas-

tecnologias da perversão social em sua materialização-discursividade do quadro conspiracionista-paranoide (CALLIGARIS, 2022).

A percepção das operações conspiracionistas-paranoicas que se fazem articuladamente de modo reivindicativo-autopunitivo é percebida em uma terceira força motriz da conspiração: o ressentimento. Tal aspecto é sinalizado por Maria Rita Kehl (2020), em sua obra *O ressentimento*, de forma que aquele que ressente acusa o Outro da posição de responsável por sua posição de ressentido (seja para reivindicação ou para autopunição), acusa-o de seduzi-lo e fazê-lo desejar (consciente ou inconscientemente) a referida posição conspiracionista-paranoide-ressentida. Este *modus operandi* se encontra em relação ao ressentimento com o Outro, contra quem as acusações-responsabilizações são lançadas e articuladas a partir de três linhas operatórias e técnicas simbólicas: a castração, a frustração e a privação.

Vê-se que a operação de tais linhas de força se fazem pela percepção dos sujeitos conspiracionistas-paranoides já imaginarem um cenário em que os modos de vida desejados já foram perdidos, e já foram perdidos pois: i) "o objeto [de desejo] perdido nunca existiu, ele é apenas uma operação que deslocou o *infans* de posição de falo para outro" (KEHL, 2020, p. 43, grifos da autora), ou seja, o mundo fantasiado nunca se sustentou como ideal, sempre o sujeito encontrou-se castrado deste, mas em sua conspiração imaginou-o como vivido e possivelmente perdido no presente por causa do Outro; ii) a perda potencial-realizada é, na verdade, um "dano imaginário. Perda imaginária de um objeto real [...] efetuada por um agente simbólico" (KEHL, 2020, p. 43), agente simbólico este que se busca vincular à figura do Outro para justificar o cenário e a condição frustrantes/frustrativas entre o esperado, fantasiado, vivido e real; e iii) há no processo de conspiração-ressentida o perceber que o objeto-realidade desejado "nunca existiu. É um objeto simbólico. [Assim] No ressentimento, a perda de que o sujeito se queixa [e conspira] é sentida como privação" (KEHL, 2020, p. 43), inclinados a pensar que o Outro contra quem se conspira, é responsável por privá-lo do ideal simbólico de mundo que desejariam viver.

Vê-se, assim, que o Outro se torna depósito e responsável pelos modos de vida pelo qual não se pode viver devido às dinâmicas subjetivas-sociais-ambientais que operam traços de castração, frustração e privação. Esta percepção ecológica das dinâmicas psíquicas torna-se intensificadas no cenário contemporâneo, ao passo que as operações neoliberais deslocam o sujeito a tratar esta condição-perversão não como social, mas enquanto individual (GUATTARI, 2009), ao passo que o sujeito que conspira passa a tomar a posição paranoide catalisadora de forças negativas-reativas não enquanto pauta da massa-grupo, mas de sua própria defesa – reivindicativa e autopunitiva – da posição de indivíduo.

Soma-se a isso o cenário em que há um processo de hipertrofia do Eu[15], em que os modos de subjetivação se fazem com um investimento narcísico que potencializa os mecanismos de defesa do Eu (KEHL, 2020). O referido cenário desloca os indivíduos a verem a preservação do Eu e daquilo que o mantém enquanto um compromisso pessoal e social, ao mesmo tempo que abre mão de mecanismos que colocariam o Eu em questão, como a responsabilização. Aquele que é responsável opera a culpa não apenas contra o Outro, mas ele investe o afeto da culpa para rever

[15] Compreende-se que o Eu é constituído a partir das pedagogias que o atravessam e permite a organização, plasticidade e estrutura psíquica. Ressalta-se que a reivindicação pelo Eu enquanto um referente-referência moderna e contemporânea encontra-se ancorada na noção de uma identidade possível. Trato a hipertrofia do Eu enquanto um processo de neonarcisificação em que o Eu desenvolve práticas e pensamentos com intuito a fortalecer seu Eu a ponto de não vislumbrar alteridade e o Outro no horizonte (POLIZEL, 2019c).

sua posição subjetiva. Na contemporaneidade o conspiracionista-paranoide-ressentido abre-se mão da responsabilidade que desestabilizaria o Eu, fortalecendo a percepção de si e deslocando a responsabilização ao Outro (REICH, 1988).

Este Outro, contudo, deve ser possivelmente rostificado (DELEUZE; GUATTARI, 2011; DELEUZE; GUATTARI, 1995), visível e não difuso. Este Outro deve ter o investimento discursivo que direciona a estes – seja este Outro um sujeito ou uma representação discursivo-simbólica –, como nos atenta Wilhelm Reich (1988), este Outro torna-se: o professor, os grupos minoritários LGBTs, os movimentos dos campos, negros, indígenas, ribeirinhos, feministas. Este Outro é aquele contra quem se pode lançar o "rosário de queixas e acusações" (KEHL, 2020, p. 29).

As queixas e acusações lançadas contra os Outro, no horizonte da conspiração, acionam um outra via-tecnologia de poder para sua manutenção, a de mobilizar a "personificação do abstrato do valor" (ADORNO, 2020, p. 26), ao passo que é preciso potencializar este Outro para conspirar contra estes. Digo, com isso, que há o investimento, percebido no sintoma Escola sem Partido, de projetar em grupos minoritários a serem combatidos por este agrupamento, um poder potencial "não vinculado ao corpo, mas a uma alma maligna" (ADORNO, 2020, p. 26) que este supostamente faz assombrar. E por ter este poder potencial, é passível (e desejável) de ser combatido.

A grande problemática é que o conspiracionismo em seu fundo ressentido adia a possibilidade de ação pela projeção do poder potencial do Outro e por sua falta de poder de ação – buscando assim alianças de cunho massificado, subjetivo ou de uma possível base jurídico-legislativa que lhe conferiria a verdade – (KEHL, 2020); e/ou pela própria estrutura do ressentimento que requer e se sustenta apenas em seu plano de fantasia imaginária de um ideal ascético de um possível mundo que pode(ria) vir-a-ser, mas nunca foi e talvez nunca estará no horizonte do porvir (NIETZSCHE, 2019), devido a isso operará o sacerdócio de convocar aliados a todo momento (o que apresenta uma linha problemática no que toca o negacionismo, visto que a negação tira do horizonte o potencial de promessa em seu ato de negação de aspecto de perversão social).

Apontaria, neste sentido, que há aqui um segundo tipo de ruptura com a possibilidade de uma gramática comum e, com isso, do diálogo e da educação, que opera ruptura com a posição frente ao Outro, que não pode estar no campo do comum pois é fonte de todo mal e responsável pelas contaminações que pululam para o conspiracionista.

Dissimulo: analfabetismo científico e desregulamentação neoliberal

"Resultado: os estudantes adquirem uma visão distorcida da realidade" (Movimento Escola Sem Partido)[16]

Acredito que no cerne da discussão que o sintoma Escola sem Partido nos coloca a pensar, no cenário de "pós-", há uma terceira linha operatória que rompe com o cenário da gramática comum – e das verdades, e das ciências, e dos coletivo –, de modo que no cerne do discurso do sintoma Escola sem Partido encontra-se a justificativa da defesa da educação e de uma perspectiva da realidade de modo não distorcida. Este investimento discursivo se faz por meio de um processo acusativo – e aqui vemos a relação com a negação e a conspiração – do que estaria adequado em termos de ser "doutrinatório" ou "científico-formativo".

[16] Perguntas e respostas do movimento Escola sem Partido. Disponível em: <https://escolasempartido.org/perguntas-e-respostas>. Acesso em 10 de janeiro de 2023

Assim vemos que a base argumentativa do sintoma Escola sem Partido é o processo de reinvocar ao cerne da discussão a formação baseada nas ciências e em sua suposta neutralidade na concepção da realidade. Este movimento conclama por produções de enunciados que demarcariam o que se encontra no parâmetro do que seria ou não científico. O científico para estes encontrar-se-ia no plano de uma definição negativa, sendo aquilo que não seria ideológico e doutrinatório.

Estes elementos são pontuados em toda a interface do site do movimento-programa Escola sem Partido: *Quem somos*[17]; *Programa Escola sem Partido*[18]; *Perguntas e Respostas*[19]; *Home*[20]; bem como nos projetos de lei apresentados pelos representantes do programa (BRASIL, 2019; BRASIL, 2016; BRASIL, 2015).

A estratégia da definição negativa é operar pelo dizer não, remetendo àquilo que não é ciência, para alocá-la em um espaço fantasmagórico do que seria possível de sê-la (SAFATLE, 2008). Contudo, este espaço fantasmático se expande em sua indeterminação incorpórea, multiplicando 'nãos' acusatórios, afirmando que determinados saberes 'não são científicos' e por isto são 'doutrinas'. Tal operação produz um potencial deslocamento das bases do que seriam possíveis para pensar a construção de um currículo. Aponto tal aspecto pela compreensão de que toda produção curricular requer a produção de um artefato cultural (SILVA, 2010), e todo artefato cultural não detém tecnologias dos sistemas e signos, significados, sentidos e representações fixos, logo só pode ser desenhado a partir de definições afirmativas.

Esta percepção colocaria uma primeira problemática na discursividade do sintoma Escola sem Partido, ao passo que as definições de cunho negativo não dariam subsídios para a proposição de um regime legal que influísse em normas-desenhos curriculares. Todavia, as definições negativas produzem efeitos e ressonâncias que não se comportam de modo simplório, sendo então o deslocamento mobilizado do sintoma Escola sem Partido realizado para o campo das "percepções de realidades", que deveriam supostamente ser 'não distorcidas'. O educar-se e formar-se encontrar-se-ia ligado ao modo pelo qual a realidade é percebida, contudo, encontra a problemática de quais parâmetros serão realizados para ancorar a percepção sobre o que é compreendido enquanto realidade. Ao considerar que a mobilização discursiva do sintoma Escola sem Partido enfoca-se em uma definição de apelação negativa, não há proposição de um quadro de referências que interconecte a realidade com a percepção. Há, assim, a proposição de um currículo que não se apresenta, mas faz negar as possibilidades curriculares vigentes – poderíamos dizer que opera um currículo negativo.

Vemos neste sentido um destoar do que tem sido articulado no campo das teorias dos currículos, que têm investido na lógica de proposição curricular a partir de pontos de referência que guiem a trajetória formativa dos sujeitos (SILVA, 2015; 2010). Desde os currículos tradicionais aos pós-críticos[21], o que se vê é o invocar e evocar de conceitos que permitam esta ancoragem e

[17] Disponível em: <escolasempartido.org/quem-somos/>. Acesso em 3 de fevereiro de 2023

[18] Disponível em: <escolasempartido.org/programa-escola-sem-partido/>. Acesso em 4 de fevereiro de 2023

[19] Disponível em: <escolasempartido.org/perguntas-e-respostas/>. Acesso em 4 de fevereiro de 2023

[20] Disponível em: <escolasempartido.org>. Acesso em 4 de fevereiro de 2023

[21] Compreende-se que os currículos apresentam uma historicidade a partir das questões que ocupam. Assim os currículos tradicionais ocupam-se das questões de ensino (o que ensinar? Como ensinar?) e aprendizagem (como avaliar o aprendido a partir do ensinado?); já os currículos críticos colocam em questionamento a intencionalidade e os sujeitos do processo educacional, sendo que emergem temas como identidade, ideologia, hegemonia, poder (para que e para quem ensinar?); e os currículos pós-

referenciação, para assim produzir estrutura e tecnologias dos sistemas e signos na proposição e operação curricular.

Este aspecto, contudo, não é considerado e percebido na discursividade do movimento Escola sem Partido, e pontuo isto ancorado na percepção de Friedrich Nietzsche (2019; 2016; 2012; 1974) e Michel Foucault (2016; 2010) de que toda elaboração de definição negativa busca ancorar-se no substrato do qual a negação é elaborada, sendo que o Escola sem Partido ancora tal negação na noção de ciências. Aquilo que não deve ser considerado aspecto curricular, não o deve por não ser científico, pontua o sintoma Escola sem Partido.

Este processo negativo-acusatório leva à disputa por retirar determinadas temáticas, saberes e discursividades do plano das ciências, arrastando-os para o plano da suposta ideologia. Esta disputa da localização de saberes no campo das ciências e nos espaços escolares[22], articula-se com o retirar a credibilidade destas e assim restringi-las de serem enunciadas no espaço de formação. Esta restrição de temática produz uma relação de forças que opera no caminho contrário do que o campo das educações e ensinos de ciências tem elaborado, ao passo que as nutrições de temáticas diversificadas e controvérsias são estimuladas nos currículos de cunho crítico e pós-crítico, com intuído da produção e operação de processos formativos de alfabetização científica.

Compreende-se que o processo de alfabetização científica – e/ou enculturação científica e/ou literacia científica – ganha espaço nas tentativas de demarcar as educações e ensinos a partir do significante 'ciências'. Este perspecto emerge em múltiplos modos de desenhar o processo formativo de modo que as ciências, enquanto conceito guia, norteiem aspectos de relevância para demarcar quais traços são formativos, ou não. Seu enfoque em aspecto amplo remete ao conjunto de práticas sociais, subjetivas e simbólicas que trazem à cena a compreensão dos modos de leitura e escrita de mundo a partir das bases do pensamento científico e da natureza da ciência (SASSERON, 2014).

A considerar as próprias balizas do conceito de ciência enquanto elaboração discursiva que se dá nas dinâmicas culturais (LATOUR, 2020; 2013; 2011; FOUCAULT, 2016; 2004), estes 'conjuntos de práticas' passam por demandas e efetivações que derivam da própria noção de ciência. Há derivações que recorrem ao apresentar a alfabetização científica a partir da compreensão de aspectos de três instâncias como a compreensão da natureza das ciências, dos conceitos científicos e dos impactos produzidos nas relações de ciência, tecnologia e sociedade (MILLER,1983); alinhado a isto, outras elaborações são apresentadas considerando dimensões que remetem a aspectos funcionais, conceituais, procedimentais e multidimensionais produzidos a partir dos processos de alfabetização científica (BYBEE, 1995); há ainda definições que derivam a alfabetização científica compreendendo a promoção de uma cultura científica no que toca práticas pessoais, sociais, culturais e para a humanidade a partir dos saberes científicos (DIAZ; ALONSO; MAS, 2003).

Segundo Attico Chassot (2000), as abordagens que buscam demarcar as alfabetizações científicas, ou seja, uma composição curricular que demarca habilidades, conceitos e possibilidades no que remete a um desenho a partir do conceito de ciência, se multiplicam e derivam à medida que as discussões sobre os modos de perceber, enunciar a praticar ciências se torcem e reelaboram. Vê-

críticos em sua desconfiança inserem mais questões na discussão, pois veem o emergir de discussões que tocam as diferenciações, alteridade (para que e para quem ensinar mesmo?) (SILVA, 2015).

[22] Ressalta-se que tal investidura incorre em um erro conceitual em seu princípio, haja visto que no espaço escolar os saberes científicos são compostos com outros saberes e em uma transposição didática se coloca enquanto saber escolar.

se, a exemplo, a multiplicação no campo da alfabetização científica no que toca os domínios de conceitos inicialmente, todavia emergem em sua sequência o pontuar a necessidade de os sujeitos dominarem atitudes e práticas que remeteriam às ciências – a exemplo, o desenvolvimento de habilidades para ações que remetem ao uso e desenvolvimento adequado de práticas, comunicação, interrelação, sistematização etc.

Gerárd Fourez (1999) pontua ainda o horizonte de pensar a alfabetização científica enquanto também um processo de alfabetização técnica, ao passo que os sujeitos passam a operacionalizar o uso de técnicas e tecnologias a partir de seus processos de formação, aspecto que pode remeter a um cenário da racionalidade técnica (ADORNO; HORKHEIMER, 1991) e também da própria compreensão da constituição de um horizonte de técnicas-tecnologias nas mediações dos modos de ser e de elaborações discursivas (FOUCAULT, 2004).

Há, neste sentido, ao pensar a alfabetização científica (e técnica), a compreensão de que são necessárias demarcações curriculares que emerjam do conceito de ciências e de elaborações conceituais, atitudinais e práticas. Este currículo-trajetória e tais possibilidades são colocadas à cena a partir da criação de condições no que tocam situações formativas, controvérsias-problemáticas e aspectos instrumentais (FOUREZ, 1999; CHASSOT, 2000), de modo que sem problemática-problema não é possível o desenho de um processo de alfabetização científica (de modo que a própria natureza da ciência demanda no campo de saber o encontro com o não saber e as possibilidades de elaborar sobre estes) (SASSERON, 2014). O que temos é a premissa de que é preciso um conceito produtivo-afirmativo de ciência-saber-técnica para que o currículo seja desenhado.

Vemos neste contraponto que a proposta do sintoma Escola sem Partido dá-se no sentido inverso por meio de uma tripla negativação. Nega saberes ao situarem-nos por definição negativa no espaço da ideologia – enquanto polo oposto à ciência –, ao mesmo tempo nega o conceito de ciência por compreendê-lo enquanto dado e consensual (e pela própria não apresentação do que compreendem enquanto saberes científicos e seus critérios de demarcação), e nega a possibilidade da alfabetização científica em sua natureza por retirar a problemática-problema ('temas contundentes') do cerne da questão. Assim, nega-se a própria natureza das ciências articuladas a seu potencial controverso, nega-se o próprio aspecto conceitual de ciência por não se ocupar e apresentá-lo enquanto elaboração enunciativa-discursiva, e nega-se o próprio aparato prático da ciência ao buscar cercear práticas pelo risco de uma suposta 'contaminação ideológica'.

Esta somatização social do sintoma Escola sem Partido tem-se por sua 'novidade' discursiva operar um currículo negativo. Por currículo negativo compreendemos o processo no qual as balizas do processo formativo do sujeito passam a ser influenciadas pelos investimentos do dizer 'não', não sendo assim propositivo em termos curriculares, mas ainda assim influenciando na formação dos sujeitos. Sinalizo aqui para nos atentarmos e não recairmos no risco de compreender as operações de poder em um aspecto subtrativo[23], sendo que até mesmo a operação negativa encontra em sua base a operacionalização de um poder de aspecto produtivo (FOUCAULT, 2016). Este

[23] Opto, neste manuscrito, em alguns momentos fazer referências ao poder em termos subtrativos, pois compreendemos, com Vladmir Safatle (2008a; 2008b), em sua hermenêutica para Michel Foucault (2015; 2010), que a noção psicanalítica de repressão corresponde à economia política, libidinal e discursiva, de modo a nos atos repressivos (no que toca o indivíduo e o coletivo) também operar pela noção foucaultiana de poder produtivo. Assim, nos momentos em que identifico possibilidades de confusão no que toca à compreensão de poder enquanto negativo (no sentido de subtração e propriedade, não de produção e relação), optou-se pela utilização da noção de subtração.

investimento de poder operacionalizado pelo currículo negativo do sintoma Escola sem Partido, que formatiza o que tratamos aqui por analfabetismo científico, encontra-se ancorado em três linhas de força: a neoliberalização do cenário educacional, a moralização dos fluxos discursivos e a subjetivação de desconhecedores.

Esta analítica-conceituação se dá no que remete o entendimento de que se as teorias do currículo (SILVA, 2015; 2010) recorrem à noção de afirmação-produção para quaisquer proposições curriculares, isso é dado pois o processo formativo e os modos de subjetivação se fazem a partir de regulamentação e normatização dos corpos (FOUCAULT, 2016; 2010; DELEUZE; GUATTARI, 2011), sendo a proposição e operacionalização curricular investidas sobre as possibilidades de modo de ser que tais currículos irão formar. Todavia, o sintoma Escola sem Partido, ao mobilizar um currículo negativo, circula linhas de força reativas e negativas, em seu avesso propondo um objeto regulatório, uma lei, para operacionalizar um movimento desregulativo, visto que nem os temas que compreendem enquanto controversos são citados. Este investimento desregulativo é característico da discursividade neoliberal no campo da educação, nos modos de subjetivação e no corpo social (HAN, 2018; DARDOT; LAVAL, 2016).

Não por menos os representantes e rede de aliados e difusores do sintoma Escola sem Partido encontram-se alinhados a *think tanks* de base discursiva neoliberais, sendo que utilizam da narrativa do Escola sem Partido enquanto plataforma política e sistema de creditação para ocupar-se dos espaços de representação do Estado e desregulá-lo por seu próprio fluxo de funcionamento[24] (CARVALHO; POLIZEL; MAIO, 2016; FURLAN; CARVALHO, 2020).

Esta investidura, ao desregular as políticas públicas educacionais e o caráter de um currículo afirmativo-propositivo, fazendo uso de uma curricularização negativa, esvazia na própria legislação a regulação normativa e as diretrizes que orientam o processo de formação. Este esvaziamento ocorre em natureza de três ordens de negativação-esvaziamento: i) a ocultação de termos-conceitos indicativos para o processo formativo, visto que o projeto de lei que representa o sintoma Escola sem Partido não indica orientações em termos curriculares pelo enfoque em seu caráter de negação (RATIER, 2016; CARVALHO; OLIVEIRA; MAIO, 2016)[25]; ii) o esvaziamento da premissa legal ao mesmo tempo em que se opera enquanto formulação legal em sua suposta validação jurídico-legislativo, visto que não apresenta orientações técnicas no que tange a própria possibilidade de entendimento e efetuação da lei (POLIZEL, 2019a; POLIZEL, 2019b); e iii) o apagamento e engajamento discursivo em combates a temáticas específicas, cerceando pela suposta negação-proibição as possibilidades de investimento em outras políticas públicas que independem desta (FURLAN; CARVALHO, 2020; ABECHE, 2013).

O cenário colocado produz desorientação e desregulamentação a medida que, no ocultar, desencaminha aspectos formativos (saber); no esvaziar, desconcerta as bases jurídicas; e no apagar e engajar no combate de determinadas pautas (que não são apresentadas enquanto conceito em sua propositiva legislativa, mas repetidamente citado em sua base discursiva), aturde e dilapida

[24] Ressalta-se que o investimento da desregulação de políticas públicas pelo próprio estado faz-se enquanto estratégia social-discursiva de trazer verniz de constitucionalidade e legitimidade do próprio ato de desmonte e desregulamentação (UTZ, 1977).

[25] Sabe-se que a ausência de determinações e de apresentações conceituais-temáticas, nos parâmetros das políticas públicas, generaliza o que se pode ou não fazer no entendimento da lei, ocultando também mecanismos norteadores para formação de estudantes, professores e equipe de agentes educacionais (RATIER, 2016).

os investimentos econômicos para estas. Este enquadre mostra-se como base do processo desregulativo enquanto princípio neoliberal (UTZ, 1977).

Esta neoliberalização discursiva faz-se agencia ao passo que tal investimento faz-se subjetivado pelos sujeitos – mesmo sem este firmar-se em termos jurídico-legislativos – orientando os processos educacionais e curriculares ao indivíduo. Assim, o espaço da desregulamentação não se sustenta apenas pelos processos de ocultação, esvaziamento, apagamento e engajamento negativo-reativo, ele o faz ao chamar à cena ao menos um aspecto afirmativo que se coloca a partir da negação: o foco do processo educacional.

Ao sintoma Escola sem Partido pontuar "[...] a visão [...] da realidade", enquanto premissa formativa, o mesmo dissimuladamente (BAUDRILLARD, 1981) o faz pelo desvio do processo educacional-curricular de suas bases conceituais (operação negativa) e situa "[...] a visão [...] da realidade" no sujeito que olha, no indivíduo (operação afirmativa). Ressalta-se que as denúncias-acusações lançadas pelo sintoma Escola sem Partido têm por enfoque um sujeito que supostamente foi 'doutrinado' ou que 'houve a tentativa', uma mãe ou um pai 'descontente', ou uma 'escola-professor que não manteve a neutralidade'. Esta discursividade sempre tem por enfoque dar um rosto a algum sujeito que deveria, por fantasia, orientar a educação – e não o currículo, as políticas públicas, temáticas ou ocupação de um tempo comum. A derivação da compreensão de realidade com enfoque no indivíduo, e não nas elaborações discursivas e coletivas, mostra-se característica do pensamento e dos modos de subjetivação neoliberais (FOUCAULT, 2008; DARDOT; LAVAL, 2016).

O deslocamento do coletivo para o indivíduo, contudo, como única linha operatória não ganha cenário representativo com a emergência do sintoma Escola sem Partido, em 2004, ganhando destaque apenas em meados de 2014-2015 (PENNA; SALES, 2017), sendo que outra linha de força-operação é agenciada pelo movimento: a moralização dos fluxos discursivos[26]. Digo, a apenas desregulamentação e ancoramento do espaço-tempo desregulamentado no indivíduo, não se mostram enquanto base significativa para o agenciamento de Outros ao sintoma Escola sem Partido, mas a apresentação de um enquadramento de valores (mesmo que genéricos e fantasmáticos-fantasiosos), torna possível o agenciamento de coletivos que conclamam por participar de um grupo-massa (CANETTI, 2019), sendo que estes percebem a dependência do(s) Outro(s) para compreender-se por definição negativa enquanto indivíduos – e, com isso, afirmarem-se empresários de si.

A força atrativa entre os 'indivíduos', neste sentido, é o conetivo da pressuposição de uma moralidade. Friedrich Nietzsche (2019) indica que a moralidade é colocada em cena enquanto um quadro de referências para definir a valorativa daqueles que precisam de um horizonte a seguir. Ao passo que são incapazes de criar valores, demandam de um Outro (individual ou massificado) uma cartela de valores a ser seguida, de modo que estes tornam-se escravos dos valores. A localidade que emerge para disponibilizar estes valores é o mercado, disponibilizando sua cartela de valores à disposição do indivíduo (NIETZSCHE, 2016). A aderência ao discurso via moralização faz-se, neste sentido, quando os indivíduos se mobilizam por ao menos cinco fios

[26] Tenho por hipótese que tal agenciamento deu-se por dois motivos: i) o conceito de indivíduo é um conceito vazio, de difícil visualização e elaboração para além das propagandísticas de representantes políticos da nova direita; ii) no que concerne a estrutura de pensamento conservativo-reativo nas massas, a ideia de indivíduo ou coletivo só se faz funcional ao passo que há um vínculo estético-ético em termos de pensamento, sendo estes elementos que requerem uma concepção de afeto-valor (Ó, 2010; NIETZSCHE, 1974).

condutores: i) o desejo de preservação de uma ordem vigente na qual o sujeito vê-se organizado-estruturado e sente-se por desejo preservá-lo/conservá-lo, seja por benefícios trazidos por esta ordem ao mesmo, pela reiteração do desejo pela massa em que se encontra ou pela incapacidade de reestruturação cognitiva (MANHEIM, 1986); ii) o desejo conservativo-reativo modalizado pela incapacidade de criação de valores e pela operação de apropriação e inversão de valores emergentes (minoritários) para reiteração de seus desejos e posições de sujeito (normativos e adoecidos) (NIETZSCHE, 2019); iii) o desejo de fazer-se diferente ao mesmo tempo que mantém-se na posição de normatividade, agenciando Outros pelo discurso controverso da 'inovação', 'novidade' e 'conservação', e do interesse da preservação da individualidade, liberdade e propriedade; iv) o desejo mobilizado pelo sugestionamento e/ou pela idealização de um Outro enquanto possível salvador do espaço 'desregulamentado' e 'possivelmente contaminado', em que ao mesmo tempo se vê espelhado e diferenciado (FREUD, 2017; CALLIGARIS, 2022); e/ou v) pelo desejo da reivindicação moral-normativa com face de se manter na posição de superioridade-maioridade e por meio desta posição reiterar a estrutura social que lhe legitima tal territorialidade (MISKOLCI, 2018; FOUCAULT, 1977; HAN 2017).

A esteira da moralidade dos fluxos discursivos leva então os sujeitos a deslocar sua referenciação, individualização e validação discursiva, que passa a reivindicar a si o nome da verdade, ciência e realidade, considerando como critério a moralidade. Contudo, sabemos que a moralidade, enquanto quadro de referência posto por um suposto ente (NIETZSCHE, 2019), encontra seus limites na torção da noção de valor. Para tal, a moralidade, enquanto mecanismo de adjetivação e modalização dos fluxos discursivos, é deslocada do plano do valor elaborado nas relações sociais para o plano do valor enquanto propriedade. A moralidade define a verdade e a moralidade é tornada propriedade, sendo esta propriedade reivindicada enquanto qualidade do grupo que expressa o desejo do sintoma Escola sem Partido[27]. No desejo de ser-moral, os indivíduos são agenciados e aglutinam-se ao sintoma Escola sem Partido, com intuito de reivindicar-se como indivíduos em posse-propriedade da moralidade enunciada pelo movimento e assim enquanto aqueles que 'entendem verdadeiramente', de forma não distorcida, da realidade.

Nesta autoidentificação e enunciação enquanto ser que percebe a realidade de forma não distorcida, validada pela aderência ao movimento Escola sem Partido que se coloca enquanto em posse da moralidade, o indivíduo defende a desregulamentação em nome da garantia de sua moralidade. Como negar a moralidade incorreria em negar o saber, o indivíduo em sua aderência ao Escola sem Partido precisa afirmar que sabe, pois assim ele afirma que é um ser moral – ele não pode reconhecer que não sabe.

Para Linsey MacGoey (2019), o não reconhecimento do espaço do não saber impossibilita o sujeito dos processos de questionamento, reflexão e diálogo, sendo estes os meios pelos quais o conhecer, a formação e as ciências se constroem. Este não reconhecimento do não-saber é produzido pelas relações de poder e estruturações construídas em nosso tempo, sendo que a

[27] Fernando Penna e Diogo Salles (2017) apresentam que o movimento Escola sem Partido ganha visibilidade na cena pública ao passo que investe na ocupação dos espaços políticos e na discussão direta de políticas públicas do Plano Nacional de Educação, ganhando força pelos investimentos político-empresariais e carismáticos-neopentecostais. Esta perspectiva pode ser complementada com os apontamentos feitos por Andrea Dip (2018), em que as leis contrárias às perspectivas das Teorias de Gêneros e Sexualidades encontram seu alicerce em representantes ou coligados à Bancada Evangélica. Em outros momentos, pontuei (POLIZEL, 2019a; 2019b) que o movimento Escola sem Partido só ganha representação e adesão ao passo que moraliza a discussão a partir da persecução e ocupação de temáticas morais que tocam questões de gêneros e sexualidades, visto que tais pautas já se mostravam moralizadas e catalisadoras de delírios no território nacional.

tecnologia de poder investida baliza as estratégias balizadas pela formação ignorância, tendo por sintoma o sujeito que não sabe que não sabe e, por isso, afirma saber enquanto uma propriedade de si. Os investimentos neoliberais-empresariais e moralizantes guiados pela subjetivação de um sujeito que reitera o saber, pois percebe a realidade a partir de sua moral-propriedade, instaura um novo modo de subjetivação no presente: os desconhecedores (*unknowers*).

Os desconhecedores (*unknowers*), enquanto modo de ser na contemporaneidade, têm por intencionalidade: i) ignorar fatos que lhes perturbariam sua ordem estrutural; ii) buscar oráculos-profetas-líderes para orientá-los em meio à desorientação não percebida como tal; iii) perceber-se como científica-filosoficamente superiores; iv) replicar informações com as quais teve contato de outra pessoa, sendo que é a identificação com o Outro o que valida a replicação e a informação para esta; v) negar para si e para os Outros a possibilidade de um 'não saber' sobre algo; vi) tomar a noção de saber e verdade vinculada às noções de identidade, moralidade e propriedade (MCGOEY, 2019); v) operar pelas linhas de força da negatividade e da reatividade, visto que a afirmação só pode operar no campo da reflexão e proposição (NIETZSCHE, 2017; DELEUZE, 2018); vi) reivindicar como mote de suas narrativas a noção de liberdade e crença, confundindo a noção de direito e a noção de possibilidade[28]. Assim, nos cenários da desregulamentação e desordenamento, bem como da moralização dos fluxos discursivos, os desconhecedores (*unknowers*) tornam-se figuras frequentes.

Tais elementos levam-me a pontuar que: se os desconhecedores (*unknowers*) são produzidos a partir da obstacularização das possibilidades de reflexão e do que se pode conhecer a partir do 'não saber', estes representam a antípoda e/ou a inversão do que seria um sujeito que guiasse sua formação pela alfabetização científica (CHASSOT, 2000; FOUREZ, 1999), visto que este suporia seu saber a partir da moralidade-propriedade que possuiria, não se abrindo ao processo de dúvida, investigação, elaboração e sistematização de saberes; se os desconhecedores (unknowers) emergem do processo de desregulamentação produzido por um currículo negativo de investimento neoliberal e pastoral (POLIZEL, 2019a), este mostra-se um projeto político de segunda inversão, ao passo que a alfabetização cientifica clama por uma ordenação curricular que delimite as possibilidades e produza investimento para formação com enfoque na delimitação das ciências, ao invés do esvaziamento proposto pelo sintoma Escola sem Partido; e ainda os desconhecedores (*unknowers*) são colocados em cena pelo investimento de uma terceira inversão, ao passo que se organiza a partir das noções de individualidade, moralidade e 'não saber que não se sabe', aspectos que são incompatíveis com a alfabetização científica, visto que as ciências são reconhecidas em termos de sua 'natureza' enquanto uma produção coletiva (FOUCAULT, 2016; LATOUR, 2011), ética e não moralizável[29] (DELEUZE; GUATTARI, 2010; LATOUR, 2020; 2013), sendo o espaço

[28] Os desconhecedores em sua narrativa de alinhamento neoliberal e carismático-neopentecostal reiteram a caracterização do indivíduo a partir da noção de liberdade e crença-moralidade-propriedade, sendo que isto lhes conferiria o direito de fazer-dizer o que lhe for desejado. Esta reivindicação, todavia, confunde a noção de 'direito' com a noção de possibilidade. No que refere noção de possibilidade, o indivíduo considera que poderá fazer-dizer pois isto encontra-se no horizonte do possível para este; isto, todavia`, não lhe confere o direito, visto que a noção de direito consiste em uma conceituação do campo jurídico e encontra-se interligado ao que pode ser avaliado por um processo de ordenamento jurídico do que é possível no campo da legalidade. Assim, poder fazer não quer dizer que o sujeito em sua liberdade e crença possui o direito, visto que a possibilidade de fazer-dizer pode incorrer na realização de um ato ilegal (logo que este não poderia e não teria o direito de fazê-lo).

[29] Vale ressaltar que houve inúmeras tentativas de moralização das ciências, sendo inúmeras atrocidades realizadas 'em nome da Ciência' (STENGERS, 2015), todavia no espaço-tempo de moralização as ciências perdem sua potência e passam a operar enquanto Deus. Neste caso, há grande esforço daqueles que advogam em nome da Ciência-Deus, visto que há dificuldade em sustentar tal percepção diante da morte de Deus (NIETZSCHE, 2012).

do 'não saber' e da controvérsia diante do emprego para descrição dos fenômenos perceptados-percebidos (DELEUZE; GUATTARI, 2010; LATOUR, 2020; 2013; 2011; 1997).

Neste cenário, o sintoma Escola sem Partido, ao posicionar a relação entre ciência, formação e realidade, desloca e inverte os três conceitos, empreendendo para a formação guiada pelo analfabetismo científico. Suas bases operatórias se fazem pela neoliberalização do cenário educacional, a moralização dos fluxos discursivos e a subjetivação de desconhecedores. Os efeitos desta produção se dão, assim, em duas linhas de definição: i) a definição negativa em relação a alfabetização científica a considerar que o analfabeto científico não se apropria de conceitos, procedimentos e atitudes ancoradas nas ciências; ii) a definição positiva, ao considerar que o analfabetismo científico incorre em formar sujeitos que se guiam pelas noções de individualidade, moralidade e 'não saber que não se sabe'. O sujeito do analfabetismo científico por excelência é o desconhecedor (*unknower*).

Considerações em tempos de "pós-"

Neste artigo, busquei apresentar as diferenciações de linhas operatórias do sintoma Escola sem Partido acerca do negacionismo, conspiracionismo e analfabetismo científico, pensando-os em termos filosóficos, pedagógicos e de seus impactos nas educações e no pensamento científico contemporâneo.

Percebemos, neste sentido, que o sintoma Escola sem Partido ancora-se nas interlocuções com o dizer 'não', ao passo que ao operar pelo negacionismo traça linhas de negação e perversão frente às discussões levantadas no tempo presente da possibilidade de gramáticas comuns e na própria negação da função-escola de ocupação de um tempo comum. Tal aspecto coloca em problemática as possibilidades de educações e ciências, a considerar que corrói as possibilidades de vinculação e estabelecimento de um plano coletivo-criativo.

A segunda linha operatória da qual lançamos mão identifica que há uma outra tecnologia e investimento traçado pelo e na arquitetura do sintoma Escola sem Partido: a conspiração. Conspiração esta que se coloca em estado de vigília, de modo a sempre acionar a existência de um Outro passível de contaminar o cenário social ou supostamente atentar contra o sujeito, instaurando um clima de medo, ódio, autopunição e da fantasia da reivindicação de um cenário possível. Este emprego discursivo-subjetivo coloca o Outro enquanto fonte de todo mal, tornando assim o princípio de sua investidura subjetiva a desconfiança, competição, vigilância, punição e ruptura com o Outro. Sabe-se no que confere os processos educacionais, científicos, subjetivos e sociais que o princípio é a relação frente e para com o Outro, sendo que, tomada a premissa conspiracionista, o Outro se torna objeto de ruindade e desventura, tornando as linhas de força negativa e reativa às operações subjetivas-sociais. Ainda neste sentido vê-se que a operação pela desconfiança e aversão ao Outro desloca o princípio dos laços sociais da coletividade para a competitividade, retomando de modo primevo relações de ruptura ao invés de conectividade. Vê-se que o efeito é a cisão de sistemas de vinculação que dão substrato para o fazer instaurações educacionais e científicas esvaziando tais campos de saber.

A terceira linha colocada em discussão apresenta o sintoma Escola sem Partido enquanto produtor de uma curricularização negativa, esvaziada, desregulamentada e de proliferação de desinformação e desconhecimento. Este processo que se faz em nome da ciência opera pela inversa do que os campos de saberes dos ensinos das ciências têm colocado em discussão, de

modo que ao invés de colocar em cena um currículo propositivo-afirmativo que se volta ao formar-se a partir de um suposto *ethos-kultur* científico (o que tem sido sugestionado pelos campos de pesquisa dos ensinos das ciências), opera seu inverso, esvaziando os currículos 'em nome das ciências' e balizando o saber científico pelos princípios da moralidade-propriedade-percepção. As três operações se vinculam pela lógica do 'dizer não', por meio da negação, do 'dizer não' aos vínculos, e pela operação de um currículo negativo-esvaziado. Esta propositiva corrói os laços de uma gramática comum, de vínculos possíveis e de currículos propositivos-afirmativos.

Referências

ABECHE, R. P. C. 2013. Liberalismo/neoliberalismo: modos de trabalho, modos de subjetivação. In: CANIATO, A. M. P.; ABECHE, R. P. C. (Org.). *Psicanálise, teoria crítica e cultura:* uma leitura psicopolítica da subjetividade contemporânea. Maringá: EdUEM, p. 63-76.

ADORNO, T. W.; HORKHEIMER, M. 1991. *Dialética do Esclarecimento.* Rio de Janeiro: Jorge Zahar.

ADORNO, T. W. 2020. *Aspectos do novo radicalismo de direita.* São Paulo: Editora Unesp.

AGAMBEN, G. 2004. *Estado de Exceção.* São Paulo: Boitempo.

AGAMBEN, G. 2019. *Signatura rerum:* sobre o método. São Paulo: Boitempo.

ALLOUCH, J. 1997. *Paranóia:* Marguerite ou a "Aimée" de Lacan. Rio de Janeiro, Companhia de Freud Editora.

BAUDRILLARD, J. 1981. *Simulacros e Simulação*. Lisboa: Relógio D'água.

BIRMAN, J**.** 2019. *Cartografias do avesso*: escrita, ficção e estéticas de subjetivação. São Paulo: Civilização Brasileira.

BRASIL. *Projeto de lei nº 193 de 2016 – Institui o programa Escola sem Partido.*

BRASIL. *Projeto de Lei nº 246 de 2019 – Institui o programa Escola sem Partido.*

BRASIL. *Projeto de lei nº 867 de 2015 – Institui o programa Escola sem Partido.*

BYBEE, R. W. 1995. Achieving Scientific Literacy, *The Science Teacher*, v.62, n.7, p. 28-33.

CALLIGARIS, C. 2022. *O grupo e o mal:* estudos sobre a perversão social. São Paulo: Editora Fósforo.

CANNET, E. 2019. *Massa e poder*. São Paulo: Companhia das Letras.

CARVALHO, F. A.; POLIZEL, A. L.; MAIO, E. R. 2016. Uma escola sem partido: discursividade, currículo e movimentos sociais. *Revista Semina – Ciências Humanas e sociais*, v. 37, n. 2, p. 193-210.

CHASSOT, A. 2000. *Alfabetização Científica:* Questões e Desafios para a Educação. Ijuí: Editora da Unijuí.

COMITE I. 2016. *Aos nossos amigos:* Crise e Insurreição. São Paulo: N-1 Edições.

DARDOT, P.; LAVAL, C. 2016. *A Nova Razão do Mundo:* Ensaio sobre a sociedade neoliberal. São Paulo: Editora Boitempo.

DELEUZE, G. 2018. *Nietzsche e a Filosofia.* São Paulo: N-1.

DELEUZE, G.; GUATTARI, F. 1996. *Mil platôs:* Capitalismo e esquizofrenia volume 3. Rio de Janeiro: Editora 34.

DELEUZE, G.; GUATTARI, F. 2011. *O anti-edipo:* capitalismo e esquizofrenia. São Paulo: Editora 34.

DERRIDA, J**.** 1997. *A farmácia de Platão.* São Paulo: Iluminuras.

DÍAZ, J. A. A.; ALONSO, Á. V.; MAS, M. A. M. 2003. Papel de la Educación CTS en una Alfabetización Científica y Tecnológica para todas las Personas. *Revista Electrónica de Enseñanza de las Ciencias*, v. 2, n. 2, p. 80-111.

DUNKER, C. I. L. 2003. Sobre a compreensão psicanalítica de paranóia. *Mental*, v.1, n.1, p. 23-37.

DUNKER, C. I. L. 2017. Subjetividade em tempos de pós-verdade. DUNKER, Christian (Orgs). *Ética e pós-verdade.* Porto Alegre: Dublimense, p. 9-42.

FALUDI, S. 2001. *Backlash:* o contra-ataque na guerra não declarada contra as mulheres. Rio de Janeiro: Rocco.

FERNANDES, A. H. 2001. O caso Aimée e a casualidade psíquica. *Agora*, v. 4, n. 2, p.73-87.

FOUCAULT, M. 2016. *Microfísica do poder.* 4 ed. Rio de Janeiro: Paz e Terra.

FOUCAULT, M. 2010. *A hermenêutica do sujeito:* curso dado em Collège de France (1981-1982). São Paulo: Editora WMF Martins Fonte.

FOUCAULT, M. 2015. *História da sexualidade I:* A vontade de saber. 3 ed. São Paulo: Paz e Terra.

FOUCAULT, M. 2008. *Nascimento da biopolítica***.** São Paulo: Martins Fontes.

FOUCAULT, M. 1990. Michel.Qu'est-ce que la critique? Critique et Aufklärung. *Bulletin de la Société française de philosophie*, Vol. 82, nº 2, pp. 35 - 63, avr/juin 1990 (Conferência proferida em 27 de maio de 1978).

FOUREZ, G. 1999. *Alfabetización científica y tecnológica.* Buenos Aires: Colihue.

FREUD, S. 2014. *Inibição, sintoma e angústia, O futuro de uma ilusão e outros textos (1926-1929).* São Paulo: Companhia das Letras.

FREUD, S. 1969. O Eu e o Isso. In: Freud, Sigmund. *Obras completas de Sigmund Freud – Tomo XIX*. Rio de Janeiro: Imago, p.23-89

FREUD, S. 2010a. *O mal-estar na civilização, novas conferências introdutórias à psicanálise e outros textos (1930-1936).* São Paulo: Companhia das Letras.

FREUD, S. 1988. *Obras completas:* Observações psicanalíticas sobre um caso de paranóia (dementia paranoides) (1911) & Notas sobre um caso de neurose obssessiva (1911). Buenos Ayres: Amorrotu.

FREUD, S. 2017. *Psicologia das massas e análise do eu.* Porto Alegre: LP&M.

FURLAN, C. C.; CARVALHO, F. A. 2020. Comunismo e gênero no Escola sem Partido: notas para não sucumbir a uma pedagogia fascista. *Revista FAEEBA – Educação e contemporaneidade*, v. 29, n. 58, p. 168-186.

GUATTARI, F. 2009. *As três ecologias.* Campinas: Papirus.

HAMON, R. 2020. Do delírio paranóico a reivindicação: crimes de gozo em nome do ideal. *Revista latino-americana de fundamentos de psicopatologias*, v.23, n.2, p. 291-321.

HAN, B.-C. 2018. *Psicopolítica:* Neoliberalismo e as novas técnicas de poder. Belo Horizonte:

Editora Âyiné.

KANT, I. 2018. *Crítica da razão prática.* São Paulo: Lafonte.

KEHL, M. R. 2020. *Ressentimento.* São Paulo: Boitempo.

LACAN, J. 1987. *Da psicose paranóica em suas relações com a personalidade.* Rio de Janeiro: Forense Universitária.

LACAN, J. 1985. *O Seminário, livro 3:* As psicoses. Rio de Janeiro: Jorge Zahar.

LATOUR, B. 2011. *Ciência em ação:* como seguir cientistas e engenheiros sociedade afora. São Paulo: Unesp.

LATOUR, B. 2013. *Investigación sobre los modos de existência.* Buenos Aires: Paidós.

LATOUR, B. 2020. *Diante de Gaia:* oito conferências sobre a natureza no Antropoceno. Rio de Janeiro: Editora UBU.

LATOUR, B. 1994. *Jamais fomos modernos:* ensaio de antropologia simétrica. Rio de Janeiro: Ed. 34.

LIPOVETSKY, G. 2005. *A Era do Vazio.* Barueri: Manole.

LYOTARD, J.-F. 2002. *A condição pós-moderna.* São Paulo: José Olympio.

MANNHEIM, K. 1986. "O pensamento conservador". In: MARTINS, José de Sousa (Org.). *Introdução crítica à sociologia rural.* São Paulo: HUCITEC, p.77-131.

MARX, K. 2013. *O capital:* crítica da economia política. Livro I: o processo de produção do capital [1867] São Paulo: Boitempo.

MASSCHELEIN, J; SIMONS, M. 2013. *Em defesa da escola:* uma questão pública. Belo Horizonte: Autêntica.

MCGOEY, L. 2019. *The unknowers:* how strategic ignorance rules the world. London: ZED.

MILLER, J. D. 1983. Scientific Literacy: a conceptual and empirical review. *Daedalus*, v. 112, n. 2, p. 29-48.

MISKOLCI, R. 2018; *Teoria queer:* um aprendizado pelas diferenças. Belo Horizonte: Autêntica.

NIETZSCHE, F. 2016. *Assim falou Zaratustra:* um livro para todos e para ninguém. Porto Alegre-RS: L&PM.

NIETZSCHE, F. 2019. *Genealogia da Moral:* uma polêmica. São Paulo: Companhia das Letras.

NIETZSCHE, F. 1974. *Obras incompletas.* 1. ed. São Paulo: Abril Cultural.

Ó, J. R. 2010. Para uma crítica das artes da existência e da ideia de consciência na modernidade: a problematização foucaultiana. In: VICENTINI; Paula Perin; ABRAHÃO, Maria Helena Menna Barreto (Orgs). *Sentidos, potencialidades e usos da (auto)biografia.* São Paulo: Cultura Acadêmica, p. 19-48.

PENNA, F. A.; SALLES, D. C. 2017. Dupla certidão de nascimento do Escola sem Partido: analisando as referências intelectuais de uma retórica reacionária. *In:* MUNIZ, Altemar de Costa;

LEAL, Tito Barros (Org.). *Arquivos, documentos e ensino de história:* desafios contemporâneos. Fortaleza: EdUECE, p. 13-38.

POLIZEL, A. L. 2019c. *Corpos e bio-virtualidades:* pedagogias do eu no vale dos homossexuais. 152 f. Dissertação (Mestrado em Ensino de Ciências e Educação Matemática). Universidade Estadual de Londrina, Londrina.

POLIZEL, A. L. 2019b. É possível uma educação para as sexualidades em meio ao desejo 'cidadãos de bem'?. In: MAIO, Eliane; OLIVEIRA, Marcio de. (Orgs). *Gênero, sexualidades e diferenças:* categoria de análise, (des)territórios de disputa. Maringá: Eduem, p 43-60.

POLIZEL, A. L. 2019a. Percepções do movimento escola sem partido: currículos pastorais e o professor enquanto catequista. *Revista Amazônida,* v.4, n.1, p. 01-16.

RATIER, R. 2016. 14 perguntas e respostas sobre o "Escola sem Partido". In: AÇÃO EDUCATIVA (Orgs). *A ideologia do movimento escola sem partido:* 20 autores desmontam o discurso. São Paulo: Ação Educativa, p. 29-42.

REICH, W. 1988. *A psicologia de massas do fascismo.* 2 ed. São Paulo: Martins Fontes. RICOEUR, P. 2010. *Tempo e Narrativa:* 3. o tempo narrado. São Paulo: Editora WMF Martins Fontes.

ROSE, N. 2011. *Inventando nossos selfs:* psicologia, poder e subjetividade. Petrópolis-RJ: Editora Vozes.

SAFATLE, V. P. 2008b. *Cinismo e a falência da crítica.* São Paulo: Boitempo.

SAFATLE, V. P. 2017. É racional parar de argumentar. In: DUNKER, Cristian (Orgs). *Ética e pós-verdade.* Porto Alegre: Dublinense, p. 125-136.

SAFATLE, V. P. 2018. *O circuito dos afetos:* corpos políticos, desamparo e o fim do indivíduo. Belo Horizonte: Autêntica Editora.

SAFATLE, V. P. 2008a. Por uma crítica da economia libidinal. *Revista IDE:* Psicanálise e Cultura, São Paulo, v. 31, n. 46, p. 16-26.

SAGAN, C. 2006. *O mundo assombrado pelos demônios.* São Paulo: Companhia das Letras.

SASSERON, L. H. 2014. *Alfabetização científica como objetivo do ensino de ciências.* Licenciatura em Ciências, São Paulo, Modulo 7, p. 47-57.

SILVA, T.T. 2010. *Currículo como fetiche:* a poética e a política do texto curricular. Belo Horizonte: Autêntica.

SILVA, T. T. 2015. *Documentos de Identidade:* uma introdução às teorias do currículo. Belo Horizonte: Autêntica.

TSING, A. L. 2019. *Viver nas ruínas:* paisagens multiespécies no Antropoceno. Brasília: IEB Mil Folhas.

UTZ, A. 1977. *Entre neoliberalismo y neomarxismo:* filosofia de uma vía media. Barcelona: Editorial Herder.

Educação como processo de formação humana: uma revisão em Filosofia da Educação ante a premência da utilidade, de Vicente Zatti e Marcos Sidnei Pagotto-Euzebio - São Paulo: FEUSP, 2022, 189 p.

Education as a process of human formation: a review in philosophy of education in the face of the urgency of utility, by Vicente Zatti and Marcos Sidnei Pagotto-Euzebio

Jacir Silvio Sanson Junior
Pontifícia Universidade Católica de Campinas (PUC Campinas)
jssjunior@gmx.com

Samuel Mendonça
Pontifícia Universidade Católica de Campinas (PUC Campinas)
samuelms@gmail.com

Na esteira de um amplo e aprofundado estudo histórico, Vicente Zatti e Marcos Sidnei Pagotto-Euzebio oferecem uma obra crítica e incisiva, que milita por uma demarcação firme a respeito do que nossa sociedade atual pretende, afinal de contas, com seu projeto educacional em vigor. Seu questionamento também se afigura como um posicionamento ético e político, ao declararem que

> [...] educação não tem como função primeira, transmitir conteúdos, profissionalizar, capacitar para competir no mercado, mas formar a humanidade no ser humano. Se as escolas e universidades têm uma função educativa, então elas estão implicadas com a formação humana, seu compromisso primeiro não é com o mercado (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 17).

A linha central do *Educação como Processo de Formação Humana* se desdobra da tese do professor Antônio Joaquim Severino que, em seu artigo *A busca do sentido na formação humana*, entende a educação como "processo de formação humana" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 16), sendo esse, precisamente, o sentido de educação que se destaca na literatura filosófico-pedagógica ocidental. Outra importante referência se encontra no argumento da filósofa estadunidense Martha Nussbaum que, em *Sem fins lucrativos: por que a democracia precisa das humanidades*, defende a ideia de que as humanidades são cruciais para a preservação da democracia e seu pluralismo de liberdades e direitos.

Essa visão enseja questionamentos de grande aptidão interpelativa, pois projeta uma realidade desejável, ao mesmo tempo em que torna explícita as condições que se busca superar, no momento presente. Como reinserir a educação como processo formativo na atualidade neoliberal e utilitarista? Haveria formas de conciliar os interesses estritamente técnicos do mercado em obter mão-de-obra para postos de trabalho, com a orientação preponderantemente humanista da cultura que gestou para a tradição ocidental a existência de um consolidado processo formativo? Sem esse componente fundante, é possível falar a rigor em "educação"?

Recebido em 12 de fevereiro de 2025. Aceito em 6 de outubro de 2025.

O trabalho de Zatti e Pagotto-Euzebio não se deixa conduzir por uma recapitulação historiográfica apenas, e soma esforços abertamente direcionados em favor de uma resposta humanística contundente. Os autores já haviam publicado um artigo na Revista *Ixtli* (ZATTI; PAGOTTO-EUZEBIO, 2021) com essa preocupação, porém concentrando-se nos termos *paideia, humanitas e Bildung*, que dão título aos três primeiros capítulos, respectivamente.

O Capítulo 1 esmiúça a gênese da *paideia* enquanto uma ideia originalmente aventada pelos gregos, forjada ao longo de uma história por meio dos trabalhos de Homero, Sófocles, Pitágoras, Sócrates, os sofistas, Platão, Aristóteles e Isócrates. Baseados na tese de Werner Jaeger (*Paideia: a formação do homem grego*), Zatti e Pagotto-Euzebio recapitulam o trabalho filosófico sobre a *areté* (excelência, virtude), elemento que possibilita compreender que "[...] a educação mais do que aprender um saber pronto é um exercício de intelecção", ou seja, um "cuidar da alma [*psyché*]" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 34). Os autores chamam a atenção para a figura de Isócrates, afirmando que enquanto "Platão funda sua filosofia e sua educação na noção de Verdade (*episteme*) em oposição à *dóxa* (opinião, aparência)" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 50), aquele prioriza a "[...] excelência da palavra, o *logos*, que é aquilo que distingue os homens dos animais e permite o desenvolvimento da civilização e da cultura" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 50).

No Capítulo 2, Zatti e Pagotto-Euzebio delineiam duas acepções da *humanitas* latina, especificando a *humanitas* cristã articulada pelo pensamento de Agostinho (*Confissões*) e Boécio (*Consolação da Filosofia*), e a talhada por pensadores como Cícero, Virgílio, Horácio, Tito Lívio e outros. Trata-se de duas fases distintas da *humanitas*, que os pesquisadores situam, primeiro, na síntese ciceroniana do século I a.C., seguida da ascensão do pensamento cristão nos séculos IV-V d.C. A respeito de Marco Túlio Cícero, os autores sublinham na obra *De Oratore* o esforço do cônsul romano em fazer uma releitura da *paideia* grega, esculpindo na imagem do orador ideal dois aspectos principais: a "[...] união entre a arte do discurso e o saber, entre a retórica e a filosofia" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 68). Trabalhando a eloquência como produto da união entre pensamento e palavra, Cícero sinalizava para o profundo senso prático do espírito romano, precavendo para que o cultivo da retórica não se precipitasse em especulações inócuas e excessivas, mas fosse capaz de se articular com as artes liberais e assim amarrar "[...] pensamento e ação, particularmente não dissociando a sabedoria da vida política e ética" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 69). Mesmo que sejam notáveis suas diferenças, críticas e oposições ao acervo filosófico pagão, pode-se falar apropriadamente de uma *humanitas* cristã, pois nela se preserva fundamentalmente a ideia de um processo formativo que, neste caso, configura-se como uma espécie de "caminhada para Deus" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 76): "A realização do ideal de perfeição humana, processo pelo qual o homem se tornava mais perfectivamente homem, é entendido como uma peregrinação em busca da santificação e tem em Cristo o ideal guiante da jornada" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 75).

A *Bildung* abastece as reflexões do Capítulo 3 que se volta à caracterização da educação moderna. Composto no alemão por *Bild* (contorno, imagem, forma) e *ung* (processo, modelagem), o termo é de difícil tradução para a língua portuguesa, contudo forma um conceito polissêmico que compartilha com a *Paideia* "[...] a significação de formação como 'modelagem' de acordo com uma imagem universal para a realização das potencialidades humanas" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 86). A *Bildung* recebe diferentes acentos, permitindo tratar o projeto

emancipatório iluminista – tal como Vicente Zatti (2012; 2015) se refere em outros de seus artigos de reflexão – nos termos de sua construção com base em Rousseau e Kant, de sua crise ante a desconstrução da metafísica, empreendida sob as suspeitas de Nietzsche e Heidegger, e de sua reconstrução crítica articulada por Adorno e Horkheimer (da Escola de Frankfurt). Na presente obra, os pesquisadores se concentram na "[...] ideia de formação [...] relacionada a elementos éticos e estéticos [...]" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 89), respectivamente enfatizados por Kant e Nietzsche, cada qual constituindo um momento do Iluminismo ligado à formação humanista. De todo modo, a *Bildung* constitui uma garantia de que, em pleno contexto moderno e em meio às realizações da cientificidade técnica e do progresso industrial, ainda se preserve uma referência de "[...] formação do homem enquanto homem em sua integralidade, formação do ser humano como um fim em si mesmo e não o adestramento ou capacitação em função de fins exteriores" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 85), sejam eles técnicos ou utilitários.

O Capítulo 5, neste mesmo eixo, reconstitui a noção de "*logos* comunicativo" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 145), de Habermas, no panorama de uma síntese pós-metafísica que ressalta não uma suposta e definitiva dissolução da metafísica, mas o revigoramento de uma formação educativa e cultural jamais reduzível para fins de capacitação e instrução. "A reconstrução habermasiana da racionalidade resgata o potencial de esclarecimento, bem como, a validade de referências éticas e políticas possibilitadoras da construção de um mundo livre, tolerante e democrático" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 20). De acordo com os autores, essa reconstrução "[...] reacende a possibilidade da retomada, em outro patamar, do ideal de educação como processo de formação humana [...]" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 20).

Marx também recebeu dos pesquisadores um capítulo exclusivo, o Capítulo 4, por conta de seu conceito de formação "omnilateral" expresso principalmente nos *Manuscritos Econômico-Filosóficos*. De acordo com os autores, Marx percebe um alinhamento entre a divisão dos meios de produção e uma dualidade que não só estrutura a escola do capitalismo, como aparta os trabalhadores "[...] da cultura geral e dos conhecimentos com caráter desinteressado [...]" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 136). O capitalista atua no interesse de prolongar o tempo e o rendimento do trabalho, em vista dos lucros potenciais oriundos do aumento da produção operária. Isso ocorre devido à divisão social do trabalho, na qual o capitalista desfruta do tempo livre que expropria do operário, impondo-lhe o "[...] trabalho excedente como mais-valor [...]" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 134). É essa dualidade que produz e reproduz a unilateralidade, de modo que, ao invés de se constituir uma formação que busca o desenvolvimento do ser humano em suas múltiplas dimensões – inclusive com tempo livre para dedicar-se a suas potencialidades ligadas ao intelecto (ensino), ao físico (ginástica) e também à instrução tecnológica –, forja-se no capitalismo apenas o "operário unilateral da ordem capitalista" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 122-123). A omnilateralidade, portanto relaciona o conceito de uma pedagogia marxista à tradição humanista que se expressa com a superação da divisão social engendrada pelo capitalismo fabril e seus processos de criação e acumulação do capital, vigente à época da publicação de *O Capital*, bem como na contemporaneidade em que se visualiza uma "[...] educação cada vez mais subjugada à lógica do capital, [e] cada vez mais a escola é sujeita à razão econômica" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 137).

Para além de um item formal da "conclusão", o leitor se depara com um sexto capítulo inti-

tulado "*Skholé* e a Formação Humana". Esse expediente, que não deixa de ser ousado ao estilo editorial mais convencional, é bastante significativo, pois

> A *skholé* grega inaugura o escolar como lugar de formação humana, pois possibilita um espaço-tempo igualitário de liberdade, no qual as atividades educativas podem se realizar como fazeres desinteressados que não possuem outros fins além de si mesmos. A liberação do jugo dos ditames das necessidades imediatas abre a possibilidade de uma experiência formativa livre (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 163).

A designação evoca, num só lance, a origem clássica grega e a tradição cultural que forjou uma educação para a formação humanista, bem como o engajamento na defesa e construção de espaços concretos onde cultivar esses processos. É nos termos de um espaço-tempo escolar que o projeto de uma formação humana integral se efetiva, graças a esta "[...] invenção possibilitadora da democratização do tempo livre [...]" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 163). Assim os autores não apenas encaminham seu argumento para um terreno concreto, como revisitam um debate fundamental, instigando o leitor a refletir a respeito da função e finalidade da escola. Trata-se de manter pujante a temática que Amaral (2018) destacava ao relatar a surpresa de Rancière em saber que seu artigo *École, production, égalité* seria traduzido no Brasil trinta anos após a publicação de 1988:

> [...] a escola, assim como a universidade e toda instituição que se propõe educar, deve se desvincular completamente das noções de produção, de eficiência, de hierarquia, que são próprias ao mundo empresarial. Somente assim poderemos formar seres humanos que não sejam considerados como peças dentro de uma máquina produtiva que não pertence à sociedade democrática e que possam, pelo seu pensamento e pelas suas contribuições à sociedade civil, transformar as relações no interior dos grupos e das instituições em prol de mais igualdade, justiça, solidariedade e liberdade para as gerações futuras (AMARAL, 2018, p. 670).

Zatti e Pagotto-Euzebio põem-se então a repercutir, no capítulo de encerramento, o problema quanto à escola estar ou não realmente livre "[...] das premências da vida econômica e das hierarquizações sociais [...]" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 164), dos ditames da profissionalização e do mercado, e se as circunstâncias de seu cotidiano realmente fazem dela um lugar oportuno em que se assegura um protegido "[...] tempo de formação enquanto cultivo desinteressado do espírito" (ZATTI; PAGOTTO-EUZEBIO, 2022, p. 164).

É claro que tal perspectiva incita uma discussão que atinge o âmago da legislação educacional. Se observarmos a Constituição Federal, ela estabelece que a educação deve visar "[...] ao pleno desenvolvimento da pessoa, seu preparo para o exercício da cidadania e sua qualificação para o trabalho" (BRASIL, art. 205) e, organizada por um plano nacional, conduzir à "formação para o trabalho" (BRASIL, art. 214, IV) e à "promoção humanística" (BRASIL, art. 214, V). Será que o trato síncrono dessas dimensões manifesta uma legislação educacional que hesita em dar contundente primazia à formação humana?

Zatti e Pagotto-Euzebio (2022) levantam uma bandeira em cujo mastro tantos outros também deixaram suas digitais. Ao discutir as relações e atritos entre a escola tradicional e os meios de comunicação emergentes na década de 1970 – aos quais designava "escola paralela" –, Porcher (1976) traçava duas opções de encaminhamento: "Definir os objetivos educativos em função das necessidades globais da sociedade incluindo, obviamente, o ponto de vista estritamente econômico" (PORCHER, 1976, p. 112, tradução nossa), ou definir "[...] a educação como um fim em si mesma" (PORCHER, 1976, p. 112, tradução nossa).

O prof. Serafim de Oliveira, que nas décadas de 80 e 90 se dedicou à documentação, estudo e difusão do pensamento de Paulo Freire (GADOTTI, 1996; PADILHA, 1996), posicionava sua reflexão educacional para o mesmo ponto de convergência, afirmando que "[...] quase todas as

teorias educacionais contemporâneas se acham centradas no treinamento e na formação do homem adjetivo, relegando a planos inferiores a formação do homem substantivo" (OLIVEIRA, 2005, p. 90). Referenciando-se com Ortega y Gasset e Jacques Maritain, postulava:

> A reflexão filosófica sobre a prática educacional visa, primordialmente, à formação do homem. No entanto, os recentes estudos sobre o homem tendem frequentemente a enfatizar mais o seu aspecto adjetivo em detrimento do aspecto substantivo, o que torna inviável uma análise global do ser humano. Esse fato, lamentável em outras áreas de conhecimento, é extremamente catastrófico para a filosofia da educação, uma vez que, dentro desse quadro de reflexão, o indivíduo pode ser treinado para tornar-se um excelente especialista sem receber, todavia, formação alguma para torná-lo pessoa. [...] (OLIVEIRA, 2005, p. 89).

Numa semelhante perspectiva, associaríamos a obra de Mário Vieira de Mello (1986) que, em *O Conceito de uma Educação da Cultura*, conclamava para uma educação humanística rigorosa e irredutível: "na *Paidéia* a Educação é tomada como um fim em si mesma, não como uma atividade tendo em vista um objetivo exterior a si própria" (MELLO, 1986, p. 32). Eis a chamada "concepção morfológica" que prima pela circularidade entre uma Educação que engendra Cultura e vice-versa, uma Educação que não lhe é indistinta, capaz de encontrar "em si mesma sua própria razão de ser": a "Educação interiorizada como forma e não exterioridade como função". Assim o diplomata, que também foi representante brasileiro na UNESCO, colocava-se em franca oposição às concepções pedagógicas funcionais norte-americanas ("concepção funcional"), onde a Educação não seria eminentemente formativa nem estaria vinculada ao que há de mais íntimo em si mesma, sempre voltada para alguma outra coisa, sempre exteriorizada em alguma outra tarefa, sempre projetada como meio de um fim não educacional.

Esses breves exemplos acenam para possíveis cursos e palestras que, ao impulso deste e demais trabalhos de Zatti e Pagotto-Euzebio, deveriam ser ofertados ao público. Além do mais, a redação exibe uma habilidosa capacidade de síntese, com retomadas que instigam para sua pronta leitura. Por mais que seus capítulos sejam anunciados com termos estrangeiros, não estamos lidando com um texto inacessível, nem cuja erudição iniba apreender o objeto de sua análise.

É especialmente relevante sublinhar o fato de se tratar de uma obra em coautoria, resultado de uma cooperação em que se destacam as circunstâncias favoráveis de um ambiente de estágio pós-doutoral, conforme os pesquisadores assinalam em uma nota de agradecimento. Evidencia-se o esforço de tornar convergente um plano de trabalho oriundo de diferentes trajetórias acadêmicas, que se faz tangível em cada capítulo. Caso em que Vicente Zatti faz ressoar algumas de suas reflexões publicadas em periódicos, nas quais problematiza a modernidade técnico-científica e a autonomia da racionalidade crítica e dialógica, o desenvolvimento tecnológico e o desenvolvimento humano, por meio de filósofos como Kant, Nietzsche, Heidegger, Habermas e outros. Caso em que Marcos Sidnei Pagotto-Euzebio desenvolve uma parcela de seu currículo em estudos clássicos e filosofia antiga, indicando nessa linha um de seus trabalhos mais relevantes, *Introdução à Filosofia da Educação: Uma Tradição Literária* (EDUSP), este escrito em parceria com o também professor da Faculdade de Educação da USP, Rogério de Almeida.

A obra é disponibilizada gratuitamente como E-book no formato pdf (1312 Kb), no Portal de Livros Abertos da USP (https://www.livrosabertos.sibi.usp.br/portaldelivrosUSP/catalog/book/767), tendo alcançado mais de 5.600 downloads entre maio de 2023 e janeiro de 2025, numa média mensal de 245 acessos. A 2ª edição foi lançada em 2023 pela Editora CRV, de Curitiba, e está disponível também em formato impresso.

Referências

AMARAL, A. G. Q. R. do. 2018. Jacques Rancière: escola, produção, igualdade. *Pro-Posições*, Campinas, v. 29, n. 3, p. 669-686, set./dez. Disponibilidade: https://doi.org/10.1590/1980-6248-2018-0121. Acesso: 16 set. 2024.

BRASIL. 2022. *Constituição (1988):* Constituição da República Federativa do Brasil (texto constitucional promulgado em 5 de outubro de 1988, compilado até a Emenda Constitucional nº 116/2022). Brasília: Senado Federal.

GADOTTI, M. 1996. Apresentação. In: GADOTTI, M. (org.). *Paulo Freire:* uma biobibliografia. São Paulo: Cortez, p. 19-24. Disponibilidade: https://acervoapi.paulofreire.org/server/api/core/bitstreams/010c2d36-b5ef-446b-8234-c4b4b806d0e5/content. Acesso: 4 mar. 2024.

MELLO, M. V. de. 1986. *O conceito de uma educação da cultura:* com referência ao estetismo e à criação de um espírito ético no Brasil. Rio de Janeiro: Paz e Terra.

OLIVEIRA, A. S. de. 2005. Filosofia e Educação. In: OLIVEIRA, A. S. de et al. *Introdução ao pensamento filosófico*. 8. ed. São Paulo: Loyola, p. 89-118.

PADILHA, P. R. 1996. Organizando uma bibliografia de Paulo Freire. In: GADOTTI, M. (org.). *Paulo Freire:* uma biobibliografia. São Paulo: Cortez, p. 660-661. Disponibilidade: https://acervoapi.paulofreire.org/server/api/core/bitstreams/010c2d36-b5ef-446b-8234-c4b4b-806d0e5/content. Acesso: 4 mar. 2024.

PORCHER, L. 1976. *La Escuela Paralela.* Buenos Aires: Editorial Kapelusz.

ZATTI, V. 2012. O Projeto educacional emancipatório moderno e a crise do esclarecimento. *Vértices*, Campos dos Goytacazes, v.14, n. 2, p. 93-116, maio/ago. Disponibilidade: https://editoraessentia.iff.edu.br/index.php/vertices/article/view/1809-2667.20120034/1487. Acesso: 8 ago. 2023.

ZATTI, V. 2015. Projeto emancipatório moderno: construção, desconstrução e reconstrução. *Pensando – Revista de Filosofia*, Fortaleza, v. 6, n. 11, p. 341-383. Disponibilidade: https://revistas.ufpi.br/index.php/pensando/article/view/3245. Acesso: 10 nov. 2023.

ZATTI, V.; PAGOTTO-EUZEBIO, M. S. 2021. Educação como processo de formação humana e suas raízes não utilitárias: paideia, humanitas e Bildung. *Ixtli – Revista Latinoamericana de Filosofía de la Educación*, [S.l.], v. 8, n. 16, p. 193-215, Disponibilidade: https://ixtli.org/revista/index.php/ixtli/article/view/157. Acesso: 13 out. 2023.

Nísia Floresta, by Nastassja Pugliese - Cambridge: Cambridge University Press, 2023, 53 p.

Nísia Floresta, de Nastassja Pugliese

Gisele Dalva Secco
California State University San Bernardino
giseledalvasecco@gmail.com

Nísia Floresta is a unique contribution to any field that, for different reasons, aims at recovering the works and legacies of intellectual women obliterated from the (widely Anglo-European) canonical record. Canonical records happen to be the result of historiographical operations, decisions historians of an epoch make about who shall and what shall not figure as benchmark authors and problems in the compendia with which we transfer and cultivate philosophical knowledge. Nastassja Pugliese's concise and original presentation of Nísia Floresta, the philosophical underpinnings of her writings, travels and cultural achievements in Latin American territory bring geopolitical and philosophical diversity to the collection *Elements in Women in the History of Philosophy* – at present still a predominantly Anglo European assemblage of female authors.

Aware of the need to contextualize her presentation of the Brazilian philosopher, Pugliese opens the book portraying the historical and geopolitical milieu in which Dionísia Gonçalves Pinto grew as Nísia Floresta Brasileira Augusta (at times only Nísia Floresta, at others Brasileira Augusta). The book is well structured and the historic-philosophical arguments adduced in favour of the author's interpretations are sound and convincing.

In the second section, Pugliese addresses one Floresta's role and place in the Brazilian intellectual history: the historical quarrel over the authorship of the (allegedly) first feminist book published in Brazil – *Direitos das Mulheres e as Injustiças dos Homens*, published by a young Nísia Floresta in 1832. Making good use of the best bibliography about this Brazilian translational *querelle*, the author unravels the threads connecting the book, and the fate of Floresta's reception in the literary circles of her time, to the name of Mary Wollstonecraft. Between the second and third sections, Pugliese shows how a British pamphlet written by an anonymous Sophia in 1739 (*Women not inferior to man*) was published in France in 1876 under the title *Droits des femmes et l'Injustice des Hommes*, misattributed to Misstriss Godwin (Wollstonecraft's married name) by the French editor, and translated from French into Portuguese by young Floresta as a translation of Wollstonecraft's *Vindication of the Rights of Woman* (1792). The precision of Pugliese's approach to this rather intricate history of translations, misattributions of authorship and hermeneutical challenges, projects a better set of lights on the processes of mystification and demystification of Floresta as the "Brazilian Wollstonecraft". Another consequence of Pugliese's revision of the quarrel about the authorship of *Direitos*' is a more accurate historical view of the reception of Wollstonecraft's ideas in South America. At the same time, it portrays the institutional framework in which *Direitos* appeared (of legislative discussions on the educational rights and curriculum design in Imperial Brazil). Pugliese gives Floresta what is rightfully hers: the authorship of the "Dedication", a prelude to the translation of Sophia.

Recebido em 10 de junho de 2025. Aceito em 6 de outubro de 2025.

In the following sections, the reader is introduced to some original features of Floresta's works, first in terms of the themes and problems she was interested in (the horrors of the Brazilian slave trade and the slavery culture, the rights of black and indigenous peoples and, mainly, theoretical and practical problems with and about women's education). Subsections 3.1 to 3.4 show Floresta's stylistic choices as an author of many literary forms (epic poems, educational novels, chronicles etc..) as well as the strong practical appeal of Floresta's writings. Pugliese is successful in showing the philosophical and pedagogical importance of Floresta's criticism of the colonialist cultural patterns that prevented Brazilian society, by then recently freed from the institutional tutelage of Portugal, from progressing towards an effective intellectual and civilisational emancipation.

The fourth part of the book marks a move towards more specific issues, starting with Floresta's ideas on the equality of men and women. The emphasis here is not only on the philosophical ideas that make Floresta an author of her own but mainly on the nuances of her dialogue with different philosophical doctrines and the illustrious philosophical methods – especially the practical Cartesianism applied to her reflections on women's rights, a methodological heritage that came to Floresta precisely by way of the Sophia pamphlet that she had earlier translated, a text largely built on Poulain de la Barre's *L'Égalité des Deux Sexes*.

Pugliese brings her own originality as a historian to the fore in the fifth section. Via a careful analysis of Floresta's chosen texts, the author scrutinizes the subliminal operation of an interpretative principle throughout Floresta's writings, a principle she dubbed "The Colonialist Principle". For Pugliese, Floresta's reflections on the situation of black enslaved people, indigenous peoples and women in Brazil are unified by their submission to the authority of tyrants, who arbitrarily oppress these groups on a variety of levels. This is the occasion for the author to deepen the examination of a fundamental thesis of Floresta's *Opúsculo humanitário*, namely, that the civilizational index of a nation is measured by the quality of its women's access to an adequate, egalitarian education. For Floresta, the overcoming of the Brazilian educational colonial, retrograde and degrading situation for blacks, indigenous people and women, depends as much on the comparative criticism to which he dedicated her entire life (Floresta's travel diaries are an unexplored source of ethnographical evidence for many of her ideas on the education of women, for they contain detailed descriptions of care practices – from breastfeeding to the way girls are expected to dress and be educated) as on a positive, curricular and pedagogical agenda. This is precisely the subject of the sixth part of the book.

Among the numerous Florestine reflections on education, Pugliese chose to analyse, in the sixth part of the book, those on the essential role of physical health in the moral development of children and the twin need to stimulate intellectual emancipation as conditions for a real liberation from the oppressiveness of colonial corsets. The content of this section also shows the organic connections between Floresta's philosophical ideas and her pedagogical practices – after all, the philosopher was also the creator and director of a long-running school for girls in the city of Rio de Janeiro (the school operated for some 17 years and was also run by her son while Floresta travelled through Europe with her daughter). The *Colegio Augusto*'s curriculum was audacious, including much more than the domestic arts normally taught to Brazilian girls who were privileged enough to have some access to formal education: it included elementary knowledge of arithmetic, foreign languages (Greek, Latin, Italian and French), as well as geography

and history. Reading these pages, it becomes evident that Floresta's pedagogical-philosophical perspective is, on the one hand, a valuable source of historical knowledge about the philosophy made in the territories colonised by Western Europe countries. On the other hand, a source of inspiration for the field of decolonial pedagogies, so fashionable (because more than necessary) nowadays. It is also worth noting that this part of the book illustrates its good general structure, for in making explicit the conceptual relations between the forms of education (as illustration and as a form of moral dignity) in Floresta's thought, by means of her original interpretation key (the Colonialist Principle) Pugliese shows how the Brazilian philosopher exercises the Cartesianism once called "logical feminism". Contrary to some accepted criticism of Modernity and Cartesian ideas, it is precisely the mind-body dualism that allows philosophers like Nísia Floresta (and Marie de Gournay before her, to name but one) to argue for equal rights between the sexes, since the differences between males and females of the human species are merely physical, not intellectual.

The seventh and final step in this journey of discovery of Nísia Floresta's philosophy consists of a broader reflection on the allocation, or possible allocations, of her work in the canonical narratives we have inherited. The incipient nature of such reflections is justified given the aims of the collection in which the book is published. However, in the interest of doing justice to the anticolonialist spirit of this elementary work, it is worth to mention Paulo Margutti's efforts to locate Floresta's philosophical thought within the historical stratifications he suggests as a chronology of Brazilian philosophy. I refer to what Margutti calls the Enlightenment Rupture (that occurred from 1808 to 1870) and the Spiritualist Eclecticism (from 1844 to 1870) (MARGUTTI, 2020). These are valuable historiographical efforts aimed at drawing a sharper picture of the scenario in which the august Brazilian philosopher was born and within which to frame the Florestine studies to come.

References

MARGUTTI, P. 2020. *História da filosofia do Brasil (1500-hoje)*: 2ª parte: A ruptura Iluminista. São Paulo: Edições Loyola.

PUGLIESE, N. 2023. *Nísia Floresta*. Cambridge: Cambridge University Press, 2023.

A *Retórica* de Aristóteles: um guia para estudantes

Aristotle's Rhetoric: a guide to the scholarship[1][2]

Luísa Madeira Mariano Leão
Universidade Federal do Rio Grande do Sul (UFRGS)
luisa.13.leao@gmail.com

Wladimir Barreto Lisboa
Universidade Federal do Rio Grande do Sul (UFRGS)
wblisboa@gmail.com

Este guia para o estudante destina-se aos alunos iniciando um estudo sistemático da *Retórica* de Aristóteles. Divide os oitenta e cinco livros e artigos nele analisados em três grandes grupos: os que lidam com o texto da *Retórica*, os preocupados com seu contexto político e intelectual e, por fim, os que discutem o significado de conceitos importantes dentro da obra.

Como afirma nosso título, este ensaio é um guia destinado a ajudar alunos durante o início de seu estudo sistemático da *Retórica*. Este guia é também uma taxonomia, e as taxonomias nunca são neutras. Parece que nosso esquema deu preferência a mais artigos do que livros, provavelmente porque os artigos, com unicidade em seu propósito, se acomodam mais facilmente nos sucessivos encaixes do nosso sistema. Certamente, a discussão de George A. Kennedy (1963) sobre a *Retórica* e seu tratamento nos livros de William M. A. Grimaldi (1972), Larry Arnhart (1981), Thomas Farrell (1993) e Eugene Garver (1994) merecem mais do que a menção passageira que recebem. Nós também favorecemos trabalhos recentes em detrimento de trabalhos mais antigos, o que parece razoável, mas talvez tenha dado destaque excessivo aos trabalhos dos filósofos incluídos em duas coleções recentes: *Aristotle's "Rhetoric": Philosophical Essays*, editado por David J. Furley e Alexander Nehamas (1994) e "*Essays on Aristotle's Rhetoric*", editado por Amélie Oksenberg Rorty (1996).

Nosso esquema possui três partes básicas - texto, contexto e conceitos - bem como uma nota sobre os usos da retórica na produção escrita. A primeira parte, "Texto", analisa as edições, traduções e referências básicas que se encontram disponíveis e discute o problema da aparente falta de unidade na *Retórica*. A segunda seção contempla obras sobre os contextos retórico, político, teórico, canônico e intelectual da *Retórica*. A motivação para grande parte do trabalho analisado neste ponto é abordar a suposta incoerência discutida na primeira seção. A terceira parte encarrega-se dos trabalhos que tratam de conceitos-chave: as provas (*pisteis*), o entimema, o exemplo (*paradeigma*), os tópicos (*topoi*) e o estilo (*lexis*), especialmente a metáfora.

Texto: Edições, Traduções e Referências Básicas

A melhor edição do texto grego é *Aristotelis "Ars Rhetorica"* (1976) de Rudolf Kassel. Há duas

[1] Por Arthur E. Walzer, Michael Tiffany, and Alan G. Gross, publicado originalmente em *Rereading Aristotle's Rhetoric*. Carbondale: Southern Illinois University Press, 2000.

[2] Texto traduzido por Luísa Madeira Mariano Leão e revisado por Wladimir Barreto Lisboa.

Recebido em 26 de abril de 2025. Aceito em 28 de agosto de 2025.

traduções recentes em inglês: George A. Kennedy, *Aristotle on Rhetoric: A Theory of Civic Discourse* (1991) e H. C. Lawson-Tancred, *The Art of Rhetoric* (1991). Há consenso geral entre os contribuidores deste volume de que a tradução de Kennedy deve ser preferida, particularmente em virtude de suas notas úteis, apêndices e de seu glossário em grego. A tradução de Freese (edição da *Loeb Library*, 1926) também é proveitosa e tem a vantagem de ter o texto em grego na página oposta, muito embora a edição na qual se embasou, a Bekker, esteja desatualizada. Também é confiável a tradução de Rhys Roberts, especialmente na versão reeditada por Barnes (1984), com correções. A antologia popular de Bizzell e Herzberg (1990) reproduz excertos de Rhys Roberts em sua redação original.

Posto que o texto é tão controverso, é importante conhecer sua história. Esta pode ser investigada em *A History of Aristotle's "Rhetoric" with a Bibliography of Early Printings* (1989), o livro meticuloso, ricamente ilustrado e de fácil compreensão de Paul Brandes. Outro relato muito claro da transmissão de manuscritos gregos é *Scribes and Scholars: A Guide to Transmission of Greek and Latin Literature* (1991) de L. D. Reynolds e N. G. Wilson. É raro que se perceba que o manuscrito mais antigo da *Retórica* é mais de mil anos posterior à sua concepção.

Citações da *Retórica* e de outras obras de Aristóteles são convencionalmente indicadas pela "numeração (ou paginação) Bekker". A linha de abertura da *Retórica*, por exemplo, é 1354a. Como Brandes explica, Immanuel Bekker (1785–1871) selecionou, dos quatrocentos ou mais manuscritos que examinou, os cem documentos que julgou mais confiáveis e os reuniu em um texto em grego de dois volumes publicado em 1831. Além de numerar as páginas desse texto, Bekker dividiu cada página em colunas. Assim, 1354a refere-se à página 1354, coluna da esquerda, da edição de dois volumes de Bekker das obras de Aristóteles (Brandes, 162-63). O número que às vezes segue uma referência - 1354a10, por exemplo - refere-se ao número da linha na coluna. Ainda que o texto de Bekker não seja mais considerado como o mais habilitado (foi superado pelo de Kassel), seu sistema de referências continua a ser o padrão.

Para obter informações sobre textos em grego antigo, estudantes devem consultar o *Thesaurus Linguae Graecae: Canon of Greek Authors and Works*, editado por Luci Berkowitz e Karl A. Squitier (1990), informalmente conhecido como "TLG". O melhor dicionário de grego é *A Greek-English Lexicon* (1968) de Henry George Liddell, Robert Scott e Henry Stuart Jones, informalmente chamado de "LSJ".

O conjunto de textos em grego de Aristóteles, assim como a maioria da literatura grega antiga, está disponível na Internet[3], graças aos esforços do *Projeto Perseus*[4]. Este site contém uma ampla biblioteca digital que, como indicado na página inicial, contém textos, arte, arqueologia, fontes secundárias e ferramentas de pesquisa. Os textos podem ser lidos em grego ou em suas versões em inglês, também estando disponíveis muitas traduções do grego para o latim. Além disso, os textos contêm links para várias ferramentas de texto, incluindo análise morfológica de qualquer palavra, acesso direto de qualquer palavra ao seu verbete nos dicionários Liddell, Scott e Jones, citações da frequência de qualquer palavra em outros autores com links para esses locais e pes-

[3] N.T. Todos os links citados no texto original no ano 2000 foram atualizados na tradução em português de 2019. Muitos dos endereços eletrônicos originais não estão mais ativos ou já não permitem acesso completo ao material aqui mencionado. Felizmente, suas versões originais encontram-se arquivadas pela organização sem fins lucrativos Internet Archive em seu banco de dados Wayback Machine, cujo conteúdo, em sua integridade, pode ser acessado pelo site https://web.archive.org/

[4] Perseus Digital Library. Ed. Gregory R. Crane. Tufts University. http://www.perseus.tufts.edu

quisas de palavras gregas. Também estão disponíveis traduções em inglês das obras de Aristóteles, assim como de outros autores, no *Internet Classics Archive*[5], na *Library of Congress Greek and Latin Texts*[6] e na *Liberty Online*[7].

Dois excelentes comentários em inglês sobre a *Retórica* são o de Cope e Sandys sobre os três livros (1877) e o de Grimaldi sobre os dois primeiros (1980, 1988). Uma coleção recente da *Landmark Series*, editada por Richard Leo Enos e Lois Peters Agnew (1998), reimprimiu ensaios importantes sobre a *Retórica*. Outra coletânea, esta editada por Keith V. Erickson (1974), embora agora datada, permite acesso conveniente a ensaios que ainda são valiosos.

Como Aristóteles é um filósofo sistemático, conhecer pelo menos alguns de seus outros trabalhos é requisito para compreender a *Retórica*. Dois excelentes guias para sua filosofia são *Aristotle* de David Ross (1923), e *The Cambridge Companion to Aristotle* (1995), editado por Jonathan Barnes. Há também uma excelente série de coleções editadas por Jonathan Barnes, Malcolm Schofield e Richard Sorabji, dos quais *Articles on Aristotle: 1. Science* (1975) é um exemplo representativo.

Para obras suplementares sobre a *Retórica,* escritas por estudiosos em comunicação oral e retórica, consulte o Índice de Comunicação *(Communication Index)*, que também está disponível em CD (como *Comm. Search*). Para trabalhos sobre os clássicos, várias fontes estão atualmente disponíveis na Internet. *Tables of Contents of Journals of Interest to Classicists* (TOCSIN) fornece um rol de mais de 150 periódicos de clássicos[8]. O acervo é dividido de acordo com o assunto e pode ser pesquisado a partir de palavras-chave. *Gnomon* também mantém um grande banco de dados[9]. A *Bryn Mawr Classical Review*[10], ou BMCR, contém centenas de resenhas de livros a partir de 1990, que também podem ser pesquisadas por palavra-chave de acordo com o texto, autor ou título. Finalmente, a *Penn Library*[11] é um site útil que fornece *links* para clássicos e periódicos importantes, incluindo TAPA (*Transactions of the American Philological Association*), BMCR, TOCSIN, SCHOLAR (um site que provê análises de texto) e outros. Outro link na *Penn Library* leva ao *Project Muse*, que contém um mecanismo de pesquisa para listagens em periódicos além de clássicos, como ciências sociais e matemática[12]. Também oferece um link para a *Arethusa*[13], uma revista de clássicos, agora disponível online.

[5] The Internet Classics Archive by Daniel C. Stevenson. http://classics.mit.edu/.

[6] Greek and Latin Classics Texts: A Library of Congress Internet Resource Page (arquivado): https://web.archive.org/web/20011129102605/http://lcweb.loc.gov/global/classics/clastexts.html.

[7] Liberty Online © 1995, Procyon Publishing (arquivado): https://web.archive.org/web/19990220215430/http://libertyonline.hypermall.com:80/Aristotle/index.html.

[8] Tables of Contents of Journals of Interest to Classicists: http://projects.chass.utoronto.ca/amphoras/tocs.html.

[9] Gnomon Bibliographic Database: http://www.gnomon.ku-eichstaett.de/Gnomon/en/Gnomon.html.

[10] Bryn Mawr Classical Review (arquivado): https://web.archive.org/web/19990220023159/http://ccat.sas.upenn.edu/bmcr/. Bryn Mawr Classical Review (versão atual): https://bmcr.brynmawr.edu/

[11] Penn Library (arquivado): https://web.archive.org/web/19990429111137/http://www.library.upenn.edu/resources/ej/ej-classics.html.

[12] Project Muse (arquivado): https://web.archive.org/web/20000303170106/http://muse.jhu.edu:80/muse.html. Project Muse (versão atual): https://muse.jhu.edu/search?action=oa_browse.

[13] Arethusa (arquivado): https://web.archive.org/web/20011107044655/http://muse.jhu.edu/journals/arethusa/. Arethusa (versão atual): https://muse.jhu.edu/journal/14.

Contextos: Retórico e Político, Teórico, Canônico, Intelectual

Se a *Retórica* é formada por uma concepção de retórica singular e unificadora[14], seus estudiosos não concordam sobre o que a integra. Em partes do texto, especialmente no primeiro capítulo, Aristóteles parece ter a intenção de tratar a retórica filosoficamente, apresentando-se como uma alternativa aos manuais tradicionais concorrentes. Apesar disso, as seções subsequentes, principalmente o Livro III, são lidas como um manual que não hesita em censurar táticas enganosas e persuasivas. Escreve George Kennedy: "Existe, portanto, no texto como um todo, um tipo de diálogo na mente de Aristóteles entre duas visões da retórica: uma que faz fortes demandas morais e lógicas ao orador e outra que é mais orientada para o sucesso no debate".

Teóricos contemporâneos mostram-se impacientes com a explicação que a geração anterior ofereceu acerca da desarmonia no texto, esforços estes revistos por William M. A. Grimaldi (1972, 28-35). Friedrich Solmsen (1929) atribuiu as lacunas à conjuntura de elaboração da *Retórica,* composta em palestras projetadas e proferidas ao longo de um vasto espaço de tempo. Segundo esta explicação, as seções mais "idealistas" são anteriores, escritas enquanto Aristóteles estava sob a influência da Academia, já as seções mais táticas, que buscam ser moralmente neutras, constituem escritas posteriores. O texto reflete a evolução dos pontos de vista de Aristóteles, mas, como questiona Glenn Most, "se Aristóteles mudou de ideia, por que não mudou seu texto?" (1994, 188). Nem todos consideram as circunstâncias de composição do texto como explicação suficientemente satisfatória. Lemos Aristóteles para experimentar um intelecto da maior integridade, não deveríamos, portanto, fazer todo o esforço possível para encontrarmos unidade na Retórica?

Estudos que buscam identificar uma visão aristotélica harmônica da retórica geralmente o fazem colocando o texto em um dos seguintes quadros: (i) o quadro político e retórico, fornecido pela reflexão sobre o contexto particular em que a *Retórica* foi escrita, (ii) o quadro teórico, oferecido pela taxonomia do conhecimento de Aristóteles, (iii) o canônico, plano fornecido pelas outras obras de Aristóteles e (iv) o quadro intelectual, propiciado pela concepção da *Retórica* em diálogo com Platão, com os sofistas e com Isócrates.

Aqueles que desejam ler com um olhar mais crítico os estudos que analisamos podem consultar Michael Leff e Carol Poster. Leff (1993) analisa com perspicácia o trabalho sobre questões "meta-teóricas" (*"metatheoretical"*) que ocuparam a formação acadêmica americana na *Retórica*. Poster (1997) critica, de modo geral, intérpretes anglo-americanos por assumirem uma "hermenêutica incontroversa" (*"unproblematic hermeneutic")* que os leva a destacar insuficientemente as questões interpretativas que confrontam qualquer intérprete de textos antigos.

Contextos Retórico e Político

Dois contextos importantes nos quais a retórica deve ser encaixada são o retórico e o político. Josiah Ober a coloca em contexto político em *Mass and Elite in Democratic Athens* (1989). Por sua vez, George Kennedy a situa e contexto retórico em *The Art of Persuasion in Ancient Greece* (1963).

O *locus classicus* do argumento de que a *Retórica* é unificada se vista como um manual

[14] N.T. O texto original grafa Retórica com inicial maiúscula e em itálico sempre que remete à obra de Aristóteles (Ars Rhetorica, Ῥητορική), conservando o substantivo "retórica" com sua grafia usual.

- e mais do que isso, como um manual para políticos de elite - é *The Intention of Aristotle's Rhetoric*, de Carnes Lord (1981). Lord sustenta que Aristóteles, em harmonia com Platão, acreditava que a retórica não é uma arte ou ciência autônoma, mas que é subordinada aos objetivos da cidade-estado, sendo potencialmente um instrumento para o estadista prudente e responsável. O objetivo da *Retórica* é fornecer motivos e meios para a elite política comunicar-se às massas. Aristóteles espera transformar o conceito de retórica sob os olhos de "homens políticos" para torná-la claramente subordinada a uma política filosófica e, assim, oferecer uma alternativa a tradição "isocrática" que iguala retórica e política. A intenção de Aristóteles é fazer da retórica um instrumento prático para o idealismo político. Jürgen Sprute (1994) sustenta, da mesma forma, que as aparentes discrepâncias entre os ideais do primeiro capítulo e a amoralidade do Livro III resultam do reconhecimento de Aristóteles de que, com o intuito de triunfar na arena política, ideais teriam de ser comprometidos.

Finalmente, uma abordagem semelhante é adotada e uma conclusão semelhante é alcançada por Glenn Most (1994) ao defender que Aristóteles pretende produzir um manual para as mentes filosóficas que almejavam carreiras políticas. A *Retórica 1.1* procura tornar a retórica respeitável para essa audiência filosoficamente atenta. O Livro III é um esforço para lhes dar o manual de táticas que precisam para ter sucesso na arena política.

Bernard E. Jacob (1996) explica as diferenças aparentes nos dois primeiros capítulos por entendê-los retoricamente, isto é, como tendo um objetivo estratégico centrado nos ouvintes. Jacob argumenta que no primeiro capítulo Aristóteles ofereceu uma polêmica em prol de uma visão exageradamente austera da retórica a fim de expor o absurdo desse extremo. Assim, esperava transformar uma audiência resistente a retórica em uma que estivesse aberta a nova, razoável ideia da arte oferecida por ele: uma visão que ocupava o meio termo entre os excessos de austeridade e os truques dos manuais. Uma abordagem diferente mas também focada na audiência de Aristóteles é adotada por James A. Berlin (1992), que, concentrando-se nas condições econômicas, políticas e sociais específicas em que a *Retórica* foi escrita, tenta explicar as fissuras no texto como esforços de Aristóteles para aplacar tanto os defensores da oligarquia quanto os da democracia.

Em uma pesquisa mais antiga, o provocativo Magia e Retórica na Grécia Antiga (1975) de Jacqueline de Romilly oferece um contexto retórico incomum como a explicação das intenções de Aristóteles na Retórica. De Romilly traça a história da retórica, de Górgias a Aristóteles, em relação à compreensão grega de magia. De acordo com De Romilly, Górgias reconheceu a magia e a retórica como tekhnai, ambas, portanto, atividades racionais sujeitas a manipulação intelectual. Para Platão, essa fusão e transformação representa um perigo para o Estado em razão do papel desempenhado pela retórica na política nos últimos trinta anos do século V. Por isso, Platão reduziu a retórica a um mero "jeito" (knack), um talento especial. Por outro lado, o objetivo de Aristóteles é paralelo ao de Górgias: elevar a retórica a um tekhnê. Seu trabalho, porém, é de resgatá-la, separando-a de uma vez por todas de seus compromissos sofísticos com a magia e o irracional.

Contexto Teórico

Como Kennedy observa, Aristóteles divide a atividade intelectual em três tipos básicos: (i) saberes teóricos, que, como as ciências, concretizam-se no conhecimento, (ii) artes práticas, i.e.,

política ou ética, que se concretizam no bem agir ou escolher sabiamente e (iii) artes produtivas, tal como a manufatura (*crafts*) ou a medicina, que trazem algo, como um poema, ou algum estado, como a saúde, à existência (*Rhet.* 12). Embora pareça haver valor em definir a retórica como um só tipo de atividade intelectual, a ideia de que ela seja um híbrido de dois ou três tipos logra o apoio de estudiosos formidáveis, incluindo não apenas Kennedy, mas também Richard McKeon (1971) e Michael Leff (1993).

Os estudos analisados nessa seção concordam que colocar a retórica na taxonomia de conhecimentos feita por Aristóteles ajudará a identificar suas intenções, mas que tipo (ou combinação de tipos) é característico da retórica enquanto disciplina tal como Aristóteles a entendia é, em si, matéria de debate. Para aqueles no campo da retórica, Barbara Warnick (1989) fornece uma boa introdução à teoria do conhecimento de Aristóteles. Warnick reexamina a filosofia aristotélica do conhecimento, identificando a retórica como uma arte produtiva, um *tekhnê*. Ela sustenta que, como *tekhnê*, a retórica só pode levar a resultados virtuosos quando sob a orientação de outro modo de julgamento: a *phronêsis*[15].

Assim como Warnick, outros que enfatizam o fato de Aristóteles classificar a retórica como *tekhnê* tendem a interpretar a *Retórica* de maneira limitada, a considerá-la como um acervo de técnicas moralmente neutras disponíveis aos oradores. Para Forbes Hill (1981) e Troels Engberg-Pedersen (1996), a *Retórica* é um trabalho técnico, não material. Hill afirma, em um ensaio que contesta diretamente o de Lois Self (discutido abaixo), que a *Retórica* é um manual, é uma arte metodológica, útil para gerar argumentos para ambos os lados de qualquer questão, mas subordinada à arte material da política. A *Retórica*, enquanto tal, "deve ser moralmente neutra". Engberg-Pedersen, ao passo que aceita a *Retórica* como moralmente neutra de modo intrínseco, argumenta que Aristóteles entendia a retórica dentro de contextos que faziam dela uma arte com viés em direção à "descoberta da verdade".

Joseph Dunne fornece uma discussão magistral da relação entre *phronêsis* e *tekhnê* em *Back to the Rough Ground* (1993). Atestar que a retórica é um *tekhnê* e não uma *phronêsis*, é parte do ônus do oitavo capítulo do livro *Theory, Technê e Phronêsis*. Essa visão coaduna com a de Oded Balaban (1986), que sustenta que, na qualidade de *tekhnê*, a retórica é uma *pôesis*, uma forma de atividade que é um meio para atingir um fim, não uma *práxis*, que é uma forma de atividade que é um fim em si mesma. Jan Edward Garrett (1987) explora as implicações da relação entre a *tekhnê* da retórica e arte mestre da política (*statecraft*). Para Garrett, o *tekhnê* de Aristóteles é sempre uma disposição ligada à verdade e uma causa de perfeição em objetos nos quais o esforço humano está envolvido. Em *Rhetoric Reclaimed: Aristotle and the Liberal Arts Tradition* (1998), Janet M. Atwill sustenta que Aristóteles conceptualizou a retórica como uma arte produtiva (uma *poêsis* e uma *technê*). Mas ela insiste que, ao fazer isso, Aristóteles não pretendia limitar a retórica a um papel gerencial ou instrumental. Ela examina as conotações de *"technê"* na mitologia grega e em Protágoras e Isócrates para identificar uma tradição que entende a retórica como uma arte contingente de invenção e intervenção, uma tradição que ela alega subsistir na *Retórica*.

Outros estudiosos julgam que a visão aristotélica da retórica compartilha muito com a sua compreensão de *phrônesis,* ou sabedoria prática. Eles tendem a ver a retórica como uma arte destinada a guiar a cidade-estado em direções racionais e éticas. Não há dúvida de que Aristóteles queria mover o estado nessas direções, a questão é apenas se ele via a retórica como um veículo

[15] Prudência ou sabedoria prática.

que permitisse tal movimento. A interpretação mais ampla da retórica aristotélica eleva o *status* da *Retórica* enquanto trabalho filosófico do *corpus aristotélico*, situando-a em uma posição mais crucial para a conquista dos objetivos políticos de Aristóteles.

Lois Self (1979) vê a *Retórica* como necessariamente promovendo os fins da ética aristotélica. Self argumenta que as qualidades da mente que a Ética a Nicômaco associa à sabedoria prática compartilham semelhanças significativas com as qualidades exigidas de um retórico efetivo. Christopher Lyle Johnstone (1980) vai ainda mais longe. Para ele, Aristóteles concebeu a retórica como necessária tanto para a prática da virtude, quanto para o tipo de deliberação que caracterizava o estado ideal.

Dois trabalhos recentes que seguem nessa direção são *Norms of Rhetorical Culture* (1993), de Thomas B. Farrell, e *Aristotle's "Rhetoric": An Art of Character* (1994) por Eugene Garver. O primeiro vincula estritamente a retórica à *práxis*, o segundo interpreta a retórica como um trabalho filosófico focado na *phrônesis*. Anteriormente, nessa mesma direção, o trabalho de Grimaldi, *Studies in the Philosophy of Aristotle's "Rhetoric"* (1972), foi influente, e sua interpretação foi em parte contestada por James Kinneavy (1987).

Contexto Canônico

Em vários pontos da *Retórica* Aristóteles compara a retórica a outras artes - mais frequente e notadamente à dialética e à política. A famosa linha de abertura da *Retórica* caracteriza a retórica como "contraparte [*antistrophos*] da dialética", e uma comparação igualmente famosa em 1.2.7 define retórica como "um certo tipo de ramificação [*paraphues*] dos estudos dialéticos e éticos (que é justo denominar de política)". Exatamente o que Aristóteles pretende com essas comparações não está claro. O trabalho analisado nesta seção tenta averiguar a relação da retórica com outras artes dentro do *corpus aristotélico* em um esforço para entender a natureza da retórica como Aristóteles a compreendia.

As discussões sobre a relação entre retórica e dialética têm uma longa história, que foi pesquisada por Lawrence Green (1990). Segundo Green, chegado o Renascimento, todas as posições básicas já haviam sido estruturadas para que os estudiosos desde então "fizessem pouco mais do que escolher lados". Alexandre de Afrodísias, filósofo da escola Peripatética, sustentou que no Livro I Aristóteles pretendia chamar atenção para as seguintes semelhanças entre dialética e retórica: nenhuma tem um assunto específico e ambas tratam de questões em que a verdade provável é o padrão mais elevado. As artes diferem nos seguintes aspectos: a dialética procede por perguntas e respostas em busca de uma conclusão generalizável, enquanto a retórica é muitas vezes monológica e busca respostas para questões particulares. Na tradição árabe, Averróis identificou como a diferença mais proeminente que, enquanto a dialética e a retórica lidam com posições contrárias, o objetivo da dialética é destruir uma delas, enquanto a retórica tenta manter as duas à vista. Finalmente, dentro da tradição escolástica, Egídio Romano sustentou que a retórica apelava às paixões (e também à razão), e tratava de questões morais específicas para públicos não sofisticados, enquanto a dialética apelava apenas para a razão e tratava da especulação universal entre auditores perspicazes.

Dois artigos recentes tomam a famosa linha de abertura como um convite de Aristóteles para compararmos a *Retórica* com os *Tópicos*. Jacques Brunschwig (1996) enfatiza a diferença nos métodos das duas obras. Ele identifica um método histórico na *Retórica*, onde Aristóteles corri-

ge e explora trabalhos anteriores, enquanto nos *Tópicos,* vê que ele aborda o assunto de maneira teórica e "a-histórica" (compreensível, uma vez que Aristóteles reivindica ser o criador de grande parte da arte). Em contraste, Robert Wardy (1996) destaca as semelhanças entre retórica e dialética em um ensaio que aborda o problema da ambiguidade moral da *Retórica*. Se o interesse de Aristóteles pela verdade no Livro I da Retórica parece desaparecer por trás de um interesse pela tática e pela vitória no Livro III, de acordo com Wardy, uma tensão semelhante caracteriza os *Tópicos*, que oferecem um catálogo de técnicas erísticas suspeitas para combinar com as também questionáveis técnicas da *Retórica*. Wardy insiste que sua intenção não é rebaixar os *Tópicos*, mas mostrar um processo inferencial comum às duas obras. Afirma ainda que é inválido partir da presença de táticas questionáveis à conclusão de que a *Retórica* é um texto inconsistente ou amoral.

Sally Raphael (1974) argumenta que deveríamos ler retoricamente a comparação de abertura feita por Aristóteles, que pretendia apenas evocar *Górgias*, onde a retórica é dita ser a contrapartida da culinária, e não destacar algo profundo. No fim das contas Aristóteles foi vítima de seu próprio esforço retórico dramático, pois o levou a uma comparação detalhada entre retórica e dialética que não resistirá ao escrutínio filosófico.

Vários estudos comparam a natureza do conhecimento na *Retórica* e na *Política*, muitas vezes para recuperar nossa compreensão do conhecimento retórico. Do ponto de vista do estudante de retórica, o valor desses estudos repousa sob suas definições experimentais de *endoxa*, as crenças comuns de uma cultura que formam a base das provas retóricas. John M. Cooper (1994) defende que, como *endoxa* são as bases dos primeiros princípios da dialética, a retórica é uma arte distinta e independente, o retórico está em posição de ver a verdade. C.D.C. Reeve (1996) também acaba por corrigir qualquer visão que limita *endoxa* às meias-verdades incontestadas que caracterizam um "conhecimento" geral de um assunto. *Endoxa* inclui, ele insiste, aquelas visões filosóficas que contestam as não examinadas, incluindo as de Sócrates e Platão. Por fim, Stephen Halliwell (1996) oferece uma descrição matizada de um meio termo entre a visão de Isócrates e Platão sobre a relação entre os primeiros princípios e *endoxa*, um meio termo no qual ele situa Aristóteles.

Contexto Intelectual

Muitos viram a comparação inicial da retórica com a dialética menos como um comentário sobre o relacionamento das duas artes e mais como um esforço para iniciar uma discussão com Platão, que havia, em *Górgias*, contrastado-as. Entretanto, mesmo aqueles que concordam que Aristóteles envolve Platão no tratado não necessariamente concordam acerca da extensão em que Aristóteles segue (ou não) os pontos de vista de seu mestre, ou o quão próximo (ou distante) ele está das perspectivas rivais dos sofistas e de Isócrates. Os estudos analisados nesta seção afirmam ou acolhem que uma posição aristotélica distinta e consistente faz-se visível quando a *Retórica* é posta em contraste ao pano de fundo das teorias que ela evoca.

Sobre a relação de Aristóteles com Platão, Everett Lee Hunt e William Grimaldi oferecem avaliações significativamente contrastantes. A obra de Hunt de 1961, *"Plato and Aristotle on Rhetoric and Rhetoricians"*, sustenta que Aristóteles está mais próximo dos sofistas do que de Platão, enquanto o trabalho *Studies in Aristotle's "Rhetoric"* de 1972 de Grimaldi argumenta que a *Retórica* cumpre o ideal que Platão evocava no *Fedro*.

Entre discussões recentes, Eckart Schütrumpf (1994) lê a retórica como um diálogo com Platão, mas concede menos ênfase ao *Fedro* e ao *Górgias* do que às *Leis*, onde encontra um quadro referencial que explica a "visão austera" do primeiro capítulo. Carol Poster (1997) ressalta que, na Antiguidade, as visões de Aristóteles eram consideradas muito mais próximas das de Platão do que são hoje, uma suposição que ela sustenta dever ser levada em conta em nossas interpretações da *Retórica*. Ademais, pelo pouco que sabemos das opiniões de Aristóteles sobre a retórica em sua obra publicada, por exemplo, a de Gryllus (que não sobreviveu), suas opiniões parecem consonantes com as de Platão. Lendo a *Retórica* sob a influência desse pressuposto, Poster argumenta que a apresentação de Aristóteles na retórica é de uma arte tornada necessária por um sistema político que investe autoridade demais em pessoas ignorantes demais para apreciar ou participar da dialética. Mary Margaret McCabe (1994) descreve um Aristóteles menos engajado, trazendo a *Retórica* como o esforço de Aristóteles para estabelecer uma linha sutil entre as visões opostas de Isócrates e Platão sobre a retórica. Essa perspectiva, ela afirma, coloca em foco a coerência básica da visão aristotélica. A leitura detalhada de McCabe se concentra em suas alegações de que Aristóteles sustenta, contra Platão, que a retórica é uma arte e, contra Isócrates, que a retórica é uma arte limitada e altamente contextualizada. Suas conclusões não são diferentes da leitura anterior de John T. Gage (1984) no contexto da pedagogia da composição.

Os artigos sobre "retórica dialética" (*dialectical rhetoric*) e "retórica retórica" (*rhetorical rethoric*) parecem, indiretamente, refletir a mesma tensão entre o Aristóteles "sofista" e "platônico". Os termos são de Carl Holmberg (1997) que contrasta uma "retórica dialética" derivada de Platão com a "retórica retórica" de Aristóteles. O objetivo da "retórica da retórica" (i.e., a de Aristóteles) não é persuasão em nome da verdade ou, de fato, persuasão de modo algum. De acordo com Holmberg, o objetivo é tornar a audiência ciente das diferentes visões da realidade que os outros defendem, e fazê-la perceber como essas visões são semelhantes e viáveis. Porém, Robert Gaines (1986) alega que a interpretação de Holmberg depende de uma tradução incorreta de linhas cruciais, e que a descontinuidade radical que Holmberg alega não está no texto. Gaines argui que Holmberg ignora os evidentes esforços de Aristóteles para alinhar a retórica à dialética. Scott Consigny (1989) afirma que essas duas perspectivas são limitadas e que Aristóteles, na verdade, apresenta uma terceira visão - uma retórica que permite ao retórico discernir os elementos persuasivos mantendo-se, ainda assim, livre de compromissos ontológicos.

Conceitos

Pisteis: os modos de prova ou persuasão

No Livro I, capítulo 2, da *Retórica*, Aristóteles identifica três modos de "prova artística" (*pístis*): *êthos* (caráter), "*pathos*" (emoção), *logos* (razão). Sua inclusão de *êthos* e *pathos* neste segundo capítulo tem causado certa consternação, pois, no capítulo anterior (1.1), Aristóteles expressou suspeitas em relação a apelos ao juiz e propôs limitar a retórica a argumentos lógicos. As seções subseqüentes parecem consistentes com 1.2: na maior parte da *Retórica*, Aristóteles trata *êthos* e *pathos* como partes legítimas da arte retórica. Embora exista um consenso geral de que os apelos do caráter se originam no interlocutor, que apelos à emoção se originam na audiência e apelos à razão se originam no próprio argumento, a aparente inconsistência na atitude de Aristóteles levou os estudiosos a, nas palavras de Kennedy, "procurar forçar o ponto de vista do primeiro capítulo em conformidade com o que se segue, fazendo distinções muito cuidadosas

sobre o que Aristóteles está dizendo" (*Retórica,* 27). Muitos dos estudos revisados nesta seção analisam cada um dos três tipos de provas artísticas, assumindo que nosso entendimento da relação entre os três se beneficiará de um entendimento mais profundo de cada um. A discussão opera em grande parte no campo da psicologia filosófica, onde, apesar das áreas de sobreposição, as investigações sobre *êthos* tendem a destacar o significado da *phronêsis*, aquelas sobre o *pathos* tendem se concentrar em elementos da cognição, enquanto as sobre *logos* consideram se a centralidade do entimema pode significar que Aristóteles o considerou o veículo para todos os três recursos- emocional e ético, e lógico.

Êthos. Aristóteles descreve *êthos* como um complexo tripartido constituído por *phronêsis* (sabedoria prática), *aretê* (virtude moral) e *eunoia* (boa vontade). O estudo sobre a natureza de *êthos* requer uma análise conjunta de seu âmbito filológico e filosófico. As áreas principais de desacordo não abordam tanto o que *êthos* é, mas, antes, se o argumento fundamentado no caráter, amplamente concebido, foi visto por Aristóteles como moral ou amoral. As áreas específicas de interesse e, às vezes, de discórdia, incluem se Aristóteles considera os apelos ao *êthos* como necessariamente decorrentes apenas do discurso e não, também, como uma função do conhecimento prévio do público sobre o caráter do falante. Aristóteles, de fato, classifica *êthos* como uma prova artística, o que poderia sugerir que está sob o controle do orador. Conquanto, se *êthos* é só uma prova artística, não teria Aristóteles entendido-o como uma construção retórica, e, portanto, não necessariamente fundamentado no verdadeiro caráter do retórico?

Uma boa introdução ao *êthos* é o *Ethotic Argument* de Alan Brinton (1986). Brinton afirma que *êthos* é um termo moralmente carregado que se refere ao caráter pessoal. Ele adverte que alguns teóricos da fala o confundiram com *ethos*, um termo moralmente neutro que denota hábito ou costume. Arthur B. Miller (1974) apresenta uma visão contrária, argumentando que Aristóteles acredita que o hábito induz o caráter e, portanto, os dois termos compartilham "consubstancialidade básica" (309). Apesar dessa discordância, Brinton e Miller apresentam um entendimento próximo de que *êthos* e seus três constituintes têm sua origem no orador, que eles são orientados para a ação e que *êthos* é sempre aplicável quando argumentos têm importância moral. Além disso, Brinton e Miller identificam a *phrônesis* na *Retórica* como o elemento de *êthos* que governa a ação, direcionando a seleção de escolhas feitas voluntariamente a partir dos hábitos ou da educação moral. Em outras palavras, as ações de uma pessoa atestam a presença ou ausência de *phronêsis* com base em normas culturais do bem e do mal.

W. W. Fortenbaugh (1992) segue esse aspecto da *phronêsis*, e a toma como uma virtude intelectual[16], como uma faculdade orientada para os meios e, porque ações louváveis surgem dela, como prova da virtude moral (*aretê*) de uma pessoa. Fortenbaugh sustenta que Aristóteles oferece uma alternativa à apresentação de *êthos* feita pelos manuais. O caso paradigmático de Aristóteles apresenta um orador norteando um juiz na direção de um falante virtuoso, sábio e bem-intencionado. Pelo menos idealmente, *êthos* funciona não como um estímulo, e não um óbice ao discernimento. Fortenbaugh acrescenta que essa visão tripartida de *êthos* não é original de Aristóteles, mas remonta a Homero. Por sua vez, Eckart Schütrumpf (1993) rastreia mais de perto o desenvolvimento desse conceito nos antecessores de Aristóteles e descobre que o que é original em Aristóteles é a ideia de que o orador não precisa possuir, mas tão somente aparentar que possui essas três qualidades.

[16] Dianoética.

A visão de Schütrumpf parece ter mais apoio entre os teóricos: que o orador aristotélico não precisa de um *êthos* verdadeiramente seu, mas pode valer-se de um *êthos* aparente - um que reflete as circunstâncias específicas de um discurso, adaptando-se a elas e às opiniões dos ouvintes. Essa "construção" deliberada e variável do *êthos* de alguém suscitou a questão de se retórica é uma arte falsa, na qual nem a mensagem nem o mensageiro são fiáveis dentre um discurso e o seguinte. Outros autores defendem a posição de Aristóteles afirmando que *êthos*, ao ser apresentado como *phronêsis,* torna-se uma base racional de juízo mediante a qual, tanto para o orador quanto para a oratória, ações passadas orientam deliberações atuais para que optem por ações que beneficiam ambos o indivíduo e a comunidade. Note, por exemplo, o ensaio de Robert C. Rowland e Deanna E. Womack (1985), bem como o ensaio de C. Jan Swearingen e a Introdução de James S. Baumlin e Tita French Baumlin, ambos em *Ethos: New Essays in Retórica e Critical Theory* (1994), editado por Baumlin e Baumlin. Artigos que em geral encarregam-se do teor moral da retórica que têm implicações para a compreensão de *êthos* incluem os de Christopher Johnstone (1980), Forbes Hill (1981) e Troels Engberg-Pedersen (1996), discutidos anteriormente.

Pathos. A maior área de interesse que vigora entre as três provas artísticas de Aristóteles está no *pathos*, em parte porque a discussão no Livro II é a análise mais completa de Aristóteles sobre as emoções da perspectiva da psicologia filosófica. A atenção acadêmica tendeu a se concentrar em duas áreas. Os interessados de modo geral na filosofia de Aristóteles investigam se ele fornece um relato filosófico das emoções ou se, na *Retórica*, oferece uma discussão popular, adequada apenas aos propósitos do retórico praticante. A segunda questão de importância para os estudiosos é a compreensão de Aristóteles sobre a natureza das emoções - a extensão em que as emoções estão ligadas, por um lado, à cognição e, por outro, ao apetite ou desejo. O último dilema tem impactos para a nossa compreensão da moralidade ou neutralidade da visão aristotélica de retórica. Para uma excelente introdução a ambas as questões, consulte o experiente e legível *Aristotle on Emotions and Rational Persuasion* (1996) de Martha Craven Nussbaum.

Uma tradição anterior considerava a apresentação de Aristóteles das emoções na *Retórica* como um relato popular, baseado em visões geralmente aceitas. Essa tradição é criticada na *Aristotle's Rhetoric on Emotions* (1970), de W. W. Fortenbaugh. Três artigos da recente coleção editada por Amélie Oksenberg Rorty (1996) tendem a aceitar e ampliar a crítica de Fortenbaugh. John M. Cooper (1996) afirma que, embora os retóricos não exijam a compreensão científica das emoções, eles certamente precisam de conhecimento genuíno para afetar as emoções da plateia. Para Gisela Striker (1996), assim como o advogado contemporâneo, o retórico não precisa de conhecimento pleno, mas requer domínio dos "resultados baseados na teoria". Striker argumenta que Aristóteles recorreu aos melhores trabalhos produzidos na Academia, incluindo o mais importante diálogo tardio de Platão, *Filebo*, para um estudo aceitável. Mas, como Aristóteles se baseou em análises já existentes, sua apresentação carece do quadro teórico uniforme que viemos a esperar do Estagirita. Dorothea Frede (1996) faz distinções dentre as falas de Aristóteles sobre as emoções na *Retórica*. Em alguns pontos (e.g., 1.10–15), Aristóteles aborda a necessidade do retórico de discutir a plausibilidade da motivação humana - a probabilidade[17] de um conjunto específico de circunstâncias manifestar-se em uma dada ação. Aqui, o conhecimento

[17] N.T. "Probabilidade", no sentido antigo, não trata do cálculo de estatísticas dentre os resultados viáveis, mas sim da credibilidade, ou seja, da característica do que é digno de fé, do que parece veracidade, do verossímil.

popular seria suficiente, até preferível. Em outros pontos, Aristóteles fornece uma consideração sobre as emoções no contexto dos esforços do retórico para efetivamente afetar os julgamentos. Nesses casos, apenas um cômputo preciso permitiria ao retórico atingir esse objetivo.

Ainda mais atenção foi dada aos aspectos cognitivos da compreensão das emoções por Aristóteles. O ímpeto surgiu com a obra de Friedrich Solmsen, *"Aristotle and Cicero on the Orator's Playing Upon the Feelings* (1938). Solmsen observou que Aristóteles rompeu com a tradição retórica por analisar os apelos emocionais ao público, não como apropriados apenas para partes específicas de um discurso (a introdução e conclusão), mas sim como uma ocorrência constante ao longo dele. Tal onipresença pode indicar que os apelos emocionais são inerentes (não anteriores ou antecedentes) aos entimemas que constituem o argumento.

Em *Aristotle on Emotion* (1975), W. W. Fortenbaugh apresenta a análise mais abrangente do *pathê*[18] em Aristóteles. O principal intuito de Fortenbaugh é estabelecer um aspecto de um *pathê* como sua causa eficiente, que ele chama de "práticas" (e.g., raiva, medo) ou "não-práticas" (pena, indignação) em proporção à sua capacidade de invocar ação e, assim, de gerar da virtude moral. O autor ainda toma as dores de discernir as emoções e os apetites de acordo com o critério da causa eficiente, em que os apetites são causados por distúrbios fisiológicos diferentes da cognição. Grande parte dos trabalhos acadêmicos mais recentes trabalha na estrutura estabelecida por Fortenbaugh. Stephen R. Leighton (1996) concorda com as conclusões de Fortenbaugh, mas não inteiramente com seu argumento. O artigo de Leighton é útil para a investigação dos diferentes significados que propomos quando dizemos que "as emoções afetam os julgamentos" - se, por exemplo, queremos dizer que as emoções são anteriores e incapacitantes aos julgamentos ou constituintes deles. Thomas Conley (1982) afirma que *pathê,* em 2.2-11, devem ser compreendidos como elementos de um processo de raciocínio topológico, que atua ou para gerar ou dissolver emoções na audiência ou como embasamento para uma análise de causas prováveis para as ações de uma pessoa. Alan Brinton (1988) sustenta que a Ética à Nicômaco de Aristóteles relaciona virtudes a paixões e ações, e que, a partir disso, *pathê* na Retórica detém uma significância moral e racional legítima em relação à ação virtuosa e justificada.

Uma divergência importante de Fortenbaugh é *Aristotle on the Metaphysical Status of Pathê* de Amélie Oksenberg Rorty (1984), que situa a classificação de Aristóteles da resposta temporária e acidental em um *continuum*. Os estímulos variam desde as palavras de um orador que provocam uma emoção até a fome que gera um rosnado no estômago. Se a experiência emocional concretiza uma potencialidade que faz parte da natureza da pessoa, não se diz que *pathos* tenha ocorrido, defende Rorty, já que a experiência não foi a causa, mas apenas a circunstância de uma condição natural. Portanto, a definição e determinação do pathos varia com a natureza do paciente ou agente. Rorty concorda com Fortenbaugh, porém, que o significado moral do *pathê* é determinado de acordo com a resposta cognitiva da pessoa que os vivencia.

Em *The Outmoded Psychology of Aristotle's Rhetoric* (1990), Alan Brinton identifica as questões fundamentais que dividem os depreciadores positivistas daqueles que têm um interesse respeitoso no *êthos* e *pathos* aristotélicos. Ele critica aqueles no âmbito da comunicação oral que descartariam a análise de Aristóteles por não se conformar aos exames empíricos contemporâneos. Para Briton, o interesse de Aristóteles por *êthos* e *pathos* não é metafísico nem sócio-científico, mas prático e baseado numa visão cognitiva das emoções.

[18] N.T. *pathea* ou *pathê* é o plural de *Pathos*.

Logos. A interpretação do *pathos* como tendo uma dimensão cognitiva tem implicações para nossa compreensão do *logos*. Pretende Aristóteles que compreendamos *logos* como manifestados através do entimema, por exemplo, como o veículo ou depósito para apelos às emoções, ou pretende ele que o entimema se aplique apenas aos apelos argumentativos orientados para o sujeito? Seria *logos* um princípio abrangente ou uma forma discricionária de demonstração? Grimaldi (1972) urgiu que pensássemos no *logos* como todo o discurso, não como uma fonte de uma das três provas artísticas, que ele identifica como *êthos*, *pathos* e *pragma* ou assunto (66). Mais recentemente, Antoine Braet (1992) apresenta uma reflexão sobre estas questões, argumentando que a forma entimemática inclui apelos de *êthos*, *pathos* e *logos*. Outros estudos úteis para a compreensão dos *logos* são os que dizem respeito ao entimema e ao exemplo, aos quais agora nos voltamos.

Entimema e Paradeigma: os meios de persuasão

A reivindicação de Aristóteles de que sua atenção ao entimema distingue sua abordagem das rivais, acrescida de sua referência a ele como o "corpo da persuasão" (1.3.1) assegurou a importância do entimema aos estudiosos. Esta atenção tendeu a focar na caracterização de Aristóteles do entimema como um "silogismo retórico" (1.2.12-14), uma vez que os estudiosos têm discutido como o entimema difere do silogismo, discussão essa que se complicou com a percepção de que "silogismo" é um termo controverso.

Obras de James A. McBurney (1936), Nancy Harper (1973) e Eugene E. Ryan (1984) são ótimos pontos de partida para o estudo do entimema. McBurney, que contextualiza o entimema dentro do *Organon* aristotélico (seus trabalhos sobre lógica e método), entende que ele difere do silogismo apodítico menos em sua extensão (silogismos também podem suprimir uma premissa) do que pela probabilidade de suas premissas e conclusões. Alega também que Aristóteles entende o entimema não apenas como um tipo de apelo lógico, mas como apto a incorporar as três provas artísticas. O propósito de Ryan é abordar as inadequações do debate sobre o entimema em muitos comentários-padrão, os quais ele avalia. Seu método distintivo é resolver os problemas através de uma abordagem específica aos muitos exemplos de entimema que Aristóteles fornece como ilustrações. Chega à conclusão de que Aristóteles apresenta o entimema como uma forma de abordar as limitações inerentes à audiência da retórica - de que os entimemas são apropriados à retórica por serem argumentos curtos, baseados em premissas plausíveis ou prováveis que estão relacionadas com os interesses dos três tipos de retórica. Aristóteles, afirma Ryan, parte do princípio de que os entimemas seriam avaliados com base na sua capacidade de persuasão, não na sua validade formal. Nancy Harper também avança de forma indutiva ao comparar exemplos ilustrativos do silogismo e do entimema nos *Primeiros Anteriores*, bem como na *Retórica*.

Lloyd Bitzer sustenta, em um ensaio que tem sido muito influente na comunicação e composição do discurso (1959), que o entimema pode ser diferenciado do silogismo, mas não por sua natureza provável nem por sua forma "truncada", uma vez que Aristóteles afirma que alguns entimemas procedem certas premissas e que alguns silogismos são também truncados. Bitzer defende que o que distingue um entimema é o seu impacto psicológico: o entimema implica uma premissa que o orador tem razões para esperar que a audiência assuma, e assim, a audiência participa de sua própria persuasão. Embora esta análise tenha sido contestada por Conley (1984),

o parágrafo final de Bitzer no qual ele afirma que o entimema é uma criação conjunta de orador e audiência é frequentemente citado em trabalhos que defendem uma interpretação dialógica ou interativia da *Retórica* de Aristóteles. J. Scenters-Zapico (1994) reflete sobre a "compreensão entimemática" como um lugar da teoria interacional da retórica e da epistemologia do construtivismo social na *Retórica* ou em outros lugares.

Há outros pontos de vista sobre o entimema. Arthur B. Miller e John D. Bee (1972) se baseiam em argumentos etimológicos e em *De Anima* para apoiar sua visão de que o entimemas tem um forte elemento afetivo, eles afirmam que os entimemas são destinados não apenas a convencer os leitores, mas também a movê-los para a ação. Jeffrey Walker (1994) defende que o entimema retira a sua força da sua apresentação dramática mais do que de uma estrutura silogística explícita ou implícita. Walker recorre a referências de outros escritores clássicos para apoiar esta interpretação estilística, uma interpretação para a qual Conley (1984) tinha anteriormente encontrado precedente em Quintiliano. O ponto de vista de Walker parece receber apoio de M. E. Burnyeat (1994), que reitera que os silogismos em si não necessariamente evocavam para Aristóteles, a estrutura formal que hoje atribuímos a eles, e que nossa visão repousa sobre uma linha corrompida dos *Segundos Analíticos*, e que, na verdade, qualquer argumento dedutivo, não importa quão breve ou válido seja, é um silogismo.

Carol Poster (1992) traça esta história de "múltiplas considerações incomensuráveis" do entimema desde sua utilização em Homero até sua descrição na literatura acadêmica moderna. Ela aconselha que renunciemos ao esforço de chegar a um consenso entre os comentadores, muito menos a um significado definitivo, em vez disso, devemos ver "entimema" como um local de revelação de atitudes culturais em relação à retórica e à linguagem em geral. Thomas Conley (1984) considera a preocupação acadêmica com o entimema algo não inteiramente saudável. A centralidade do entimema na *Retórica* tem, segundo ele, resultado na nossa sobrevalorização da importância do argumento e da lógica e na proporcional negligência em relação ao papel que os recursos literários da linguagem desempenham na persuasão.

O outro método de prova através do *logos*, dentro do sistema aristotélico, é o exemplo ou paradigma (*paradeigma*). É também um conceito contestado. O cerne do debate entre teóricos é a afirmação de Aristóteles de que o paradigma "é raciocinar nem na relação da parte para o todo nem do todo para a parte, mas da parte a parte, semelhante para semelhante, quando duas coisas estão sob o mesmo gênero, mas uma é melhor conhecida que a outra" (1.2.19). Os intérpretes da *Retórica* não concordaram sobre o que Aristóteles pretende com esta afirmação.

Gerald Hauser (1968, 1985) e Scott Consigny (1975) assumem que a insistência de Aristóteles na ideia de que o paradigma argumenta "da parte a parte" significa que os leitores inferem com base em dois casos similares - nos termos de Hauser, que o paradigma funciona como uma "inferência não mediada" entre dois particulares (1968, 88). Não pode, então, haver qualquer implicação ou inferência de uma generalização sobre todos, ou sobre a maioria dos casos, sem borrar distinção entre exemplo (*paradeigma)* e entimema. Em artigos de resposta aos de Hauser e de Consigny, William L. Benoit (1980 e 1987) desafia esta visão. Benoit entende o raciocínio de Aristóteles de "parte para parte" como uma forma abreviada para o raciocínio "de parte para todo para parte" (1987, 264). Ele cita a análise de Aristóteles sobre a indução nos *Primeiros Analíticos*, bem como a sua abordagem do paradigma na *Retórica*, em apoio à afirmação de que uma generalização mediadora é necessária para que o paradigma funcione argumentativamente.

James C. Raymond (1984) não entra diretamente na discussão de Benoit com Hauser e Consigny, embora sua contribuição à nossa compreensão não seja minorada por isso. Raymond afirma que Aristóteles previa que o ouvinte fosse experimentar um paradigma como um padrão analógico, diacrónico. A partir de uma série de eventos de consequências conhecidas (*paradeigma,* exemplos), o leitor infere um padrão possivelmente aplicável a eventos semelhantes de consequências desconhecidas. Para Raymond, essa sua interpretação nos ajuda a entender a observação (enigmática para alguns) feita por Aristóteles de que os paradigmas são "mais apropriados para a oratória deliberativa" (3.17.5) uma vez que ela frequentemente se baseia em padrões de eventos oriundos de um passado já conhecido para prever consequências futuras.

Tópicos (topoi): retórico, dialético

Há um consenso geral de que na *Retórica* os tópicos são meios de invenção retórica, e que Aristóteles nomeia dois tipos: comum e especial. Os tópicos comuns aplicam-se a todos os gêneros de discursos retóricos, já os tópicos especiais aplicam-se apenas a gêneros específicos, tais como o epidíctico, ou a assuntos específicos, tais como a física. Para usar os exemplos de Aristóteles, o tópico do *mais* e do *menos* pode aplicar-se indiferentemente ao discurso epidíctico, deliberativo ou forense: "Se nem mesmo os deuses sabem tudo, os seres humanos dificilmente podem sabê-lo'. Pois isso equivale a dizer: 'Se algo não é o fato em um caso onde seria mais esperado, é claro que não é um fato onde seria *menos* esperado'" (2.23, grifo nosso). Por outro lado, o tópico da justiça aplica-se especificamente ao discurso forense, onde a justiça está centralmente em questão. No entanto, mesmo esta distinção não está isenta de confusões: ainda que os tópicos especiais da física não sejam aplicáveis a nenhum gênero de retórica, a justiça pode ser um tópico de discurso deliberativo ou epidíctico. Há outro problema: não está claro se os tópicos, especiais ou comuns, são materiais ou formais, ou seja, se pertencem ao domínio da semântica ou ao domínio da sintaxe, um domínio ao qual a lógica formal também pertence.

Donovan Ochs (1969) faz uma contribuição significativa para o debate quando argumenta que os tópicos na *Retórica* não constituem um sistema lógico de invenção, uma vez que deixam de fora elementos essenciais da lógica formal tais como identidade e contradição. Esta percepção também sugere a natureza aleatória das listas desorientadas de Aristóteles e, talvez, a nossa incapacidade de torná-las compreensíveis em seu agregado. Este pode ser o argumento de Michael Leff quando diz que os topos retóricos de Aristóteles são "uma noção confundida" (1983, 23).

Ochs está em terreno ainda menos seguro quando sustenta que os vinte e oito topoi em 2.23 são "padrões formais de relações existentes entre classes de termos" (422-23). Por exemplo, o primeiro destes vinte e oito é um argumento a partir dos opostos: "Se a guerra é a causa dos males presentes, as coisas deverão ser corrigidas ao fazermos a paz". Nisto, o modelo formal não determina a conclusão como faria no caso de um silogismo válido. Poder-se-ia também argumentar: "Mesmo que a guerra seja a causa dos males presentes, as coisas não podem ser corrigidas fazendo a paz". Claramente, a persuasão depende tanto do formal quanto do material: esta guerra e esta paz.

Qual é a diferença entre o topos comum e o especial? Em um importante artigo, Thomas Conley (1978) defende, usando a terminologia de *The Uses of Argument*, de Stephen Toulmin, que a diferença "não é de matéria *vs*. forma, mas sim de graus relativos de 'dependência de cam-

po/invariância de campo'". A diferença não é de tipo, mas de ênfase. Carolyn Miller (1987) estende um *insight* de Michael Leff sobre este conflito entre o formal ou inferencial e o material: ela sugere que os tópicos comuns, que enfatizam o inferencial, são um produto de necessidade pedagógica, enquanto os tópicos especiais, que enfatizam o material, são um produto da prática oratória.

Lexis ou Estilo

Os capítulos 2 a 12 do Livro III da *Retórica* focam exclusivamente na *lexis*, e por isso constituem o tratamento mais completo do tema no corpus aristotélico. O termo "*lexis*" tem ampla aplicação em Aristóteles, variando de "estilo" (*style*), sua referência mais geral, a "fonema" *(phoneme)* na *Poética* 20, para a qual Aristóteles nos remete em 3.1.9. A raiz proto-indo-europeia de "*lexis*" é ''leg-'', exatamente a mesma raiz que a dos *logos*, enquanto seu sufixo, "-sis", indica um estado, condição, qualidade, ou processo relacionado ao mesmo complexo rico de ideias indicado pelo virtualmente intraduzível *logos*.

Como resultado desta gama de significados, a relação entre *lexis* e *logos* na *Retórica* provou ser uma fonte de desacordo entre a filosofia e a retórica. No contexto desta discussão, foi identificada uma nascente filosofia aristotélica da linguagem. As observações de Aristóteles sobre a metáfora foram muito frutíferas a este respeito.

A questão de saber se Aristóteles entende ou não a *lexis* como subordinada a (ou potencialmente coextensiva com) *logos* tem confrontado os leitores por dois milênios. Aristóteles escreve em 3.2.1 que a virtude cardinal (*aretê*) da lexis como estilo é "clareza". Esta afirmação implica que a clareza não é a ausência de "estilo", e sim a sua conquista, tal como o Livro III acaba por deixar claro. Esta afirmação implicaria, então, que Aristóteles encara a linguagem e o pensamento como sendo coextensivos? Por outro lado, em outros lugares do Livro III Aristóteles parece considerar o estilo como antecedente e acessório ao pensamento. A compreensão aristotélica da relação entre linguagem e pensamento motiva Stephen Halliwell em *Style and Sense in Aristotle's Rhetoric, Bk. 3* (1983). Halliwell conclui que, embora o vocabulário analítico de Aristóteles sugira frequentemente uma divisão estilo-sentido, os detalhes da apresentação da *lexis* no Livro III não refletem uma "separação radical entre estilo e sentido", ao contrário, o Livro III transmite uma consciência das muitas maneiras pelas quais as escolhas estilísticas "podem ajudar a determinar tanto o significado quanto a força expressiva do que é transmitido pelas palavras" (66-67).

No que diz respeito à *lexis*, dentre todos os assuntos, nenhum recebeu mais atenção acadêmica do que o trato da metáfora em Aristóteles. O ensaio de Richard Moran é um bom ponto de partida. Moran (1996) observa que Aristóteles se concentra na noção de transferência inerente à metáfora, e explica as relações (por exemplo, de espécie para gênero, de gênero para gênero) no coração da análise de Aristóteles. Esta explicação é um prelúdio à sua análise da resposta psicológica que Aristóteles vê como responsável pelo impacto que a metáfora, apropriadamente utilizada, tem em contextos persuasivos. Ainda segundo Moran, o efeito da metáfora depende, para Aristóteles, do sentimento de descoberta dos ouvintes, enquanto estes se orgulham de terem percebido a semelhança pretendida.

O recente "*Aristotle on Metaphor"* (1997) de John T. Kirby é o tratamento mais abrangente até hoje, em inglês, das visões de Aristóteles sobre metáforas. Kirby coloca as observações de Aristóteles dentro de relevantes contextos: das teorias do século XX, da prática na Antiguidade anti-

ga (em Homero) e da discussão contemporânea (em Platão e Isócrates). Este enquadramento é um prelúdio ao trabalho de Kirby sobre as notas de Aristóteles, sobretudo (mas não apenas) na *Poética* e na *Retórica*. A análise de Kirby, que é informada por uma filologia sofisticada, conclui que a semiótica é a lente que melhor esclarece a compreensão de Aristóteles, e que, para este, uma metáfora eficaz acrescenta sofisticação ao discurso e produz prazer na audiência.

Muitos pesquisadores começam sua consideração sobre a metáfora comparando seu significado literal ("transpor") com sua definição em *Poética 21* (1457b6-7), onde esta ideia de transporte é dada pelo termo "*epiphora*". Paul Gordon traduz "*epiphora*" como "suplemento" em seu *The Enigma of Aristotelian Metaphor: A Deconstructive Analysis* (1990), um texto que encontra elementos do irracional e do intuitivo na teoria de Aristóteles. Em seu *Cognitive Aspects of Aristotle's Theory of Metaphor* (1984), Pierre Swiggers contesta autores que tenham atribuído a Aristóteles uma visão formalista e ornamental da metáfora como substituição verbal. Swiggers entende a contribuição de Aristóteles como uma análise de como a metáfora funciona no nível cognitivo. Baseando-se na *Poética* e na *Metafísica*, Swiggers afirma que, para Aristóteles, a metáfora funciona como uma interação orgânica e racional entre orador e ouvinte que produz primeiro o reconhecimento e, depois, o conhecimento. Em *Aristotle's Analogical Metaphor*, Steve Nimis (1988) usa uma análise marxista para investigar as implicações sociais da teoria da metáfora de Aristóteles.

Produção escrita[19] e usos da *Retórica*

Embora ensaios de retóricos em composição que abordam um ou outro conceito importante da *Retórica* tenham sido considerados em seções anteriores deste ensaio, alguns trabalhos de redatores sobre a *Retórica* têm sido mais globais, tentando ver a teoria de Aristóteles a partir de perspectivas modernistas ou pós-modernistas. A interpretação da *Retórica* é conduzida através do contexto de um argumento sobre o caráter e a direção da composição como disciplina. O ensaio de Andrea A. Lunsford e Lisa S. Ede, *On Distinctions Between Classical and Modern Rhetoric*, na influente antologia *Essays on Classical Rhetoric and Modern Discourse* (1984b), faria da *Retórica* a base fulcral dos estudos de produção escrita. As autoras defendem em seu texto que a teoria desenvolvida na *Retórica* é a teoria mais apropriada para uma disciplina comprometida com visões epistêmicas e dialógicas da retórica. Que a *Retórica* apresente uma teoria genuinamente dialógica é uma ideia contestada por Walzer (1997), que argumenta que as (des)apropriações modernas deste tipo nos impedem de ouvir a resposta genuína de Aristóteles aos problemas que a retórica monológica coloca. Em contraste com Lunsford e Ede, Jasper Neel ataca Aristóteles em *Aristotle's Voice* (1994), declarando-o responsável por muito do que atormenta particularmente a produção escrita na qualidade de disciplina e a cultura ocidental em geral. De acordo com Neel, Aristóteles valorizou um cientificismo desinteressado e objetivo que fomentou a rejeição da retórica como disciplina, enervou nosso ensino e nossa academia, e permitiu que, tanto ele quanto nós, encobríssemos uma ideologia que promove o racismo e o sexismo. Se Neel não eliminaria totalmente a *Retórica* da composição, ele pelo menos insistiria que estudiosos e professores de produção escrita assumissem uma posição crítica em relação a ela. Embora estes usos de Aristóteles sejam úteis para estimular o tipo de debate que mantém a *Retórica* viva, eles podem finalmente provar apenas que o trabalho de um gênio não se presta facilmente à polêmi-

[19] N.T. No original, *composition*.

ca. Esta lição está dentre as que Kathleen Welch oferece em *Contemporary Reception of Classical Rhetoric* (1990). Welch adverte os compositores para desconfiarem das apresentações monolíticas e tendenciosas da retórica clássica, que fizeram dela um alvo fácil para críticos.

Três livros acadêmicos também merecem ser mencionados: *Classical Rhetoric* e *Modern Student* (1965, 3rd ed., 1990) de Edward P. J. Corbett e *Rhetoric in contemporary Students* (1994) de Winifred Horner. É verdade que cada um desses livros recorre a conceitos desenvolvidos por autores clássicos anteriores e posteriores a Aristóteles. Também é verdade que cada um deles manifesta uma ênfase ligeiramente diferente dentro da rica tradição clássica: a de Crowley, por exemplo, é mais influenciada pelos Sofistas. No entanto, cada um traz inconfundivelmente a marca de Aristóteles, e todos são aquisições modernas bem-sucedidas da *Retórica*, úteis tanto para estudantes quanto para seus professores.

O racional *versus* o razoável

The rational versus the reasonable[1]

Raquel Cipriani Xavier
Universidade Federal de Santa Catarina (UFSC)
cipriani.raquel@gmail.com

1. Explicações conflitantes sobre o papel da razão na orientação da conduta. Provavelmente, poucos termos ocorrem com mais frequência na literatura da teoria moral do que "racional", "razoável" e seus antônimos. No entanto, na discussão filosófica sempre houve, e continua a haver, uma discordância considerável quanto ao que tais termos significam quando aplicados à conduta. Muitos teóricos consideram "razão" como um princípio que exige de um agente moral algum tipo de atitude utilitarista em questões de conduta: para eles, a pessoa "racional" é aquela que agirá de modo a "maximizar valores"[2]. Outros, como Kant, procuram derivar da "razão" um princípio de equidade formal, como o imperativo categórico. Outros ainda concordam com o espírito da máxima de Hume, segundo a qual "Não é contrário à razão eu preferir a destruição do mundo inteiro a um arranhão em meu dedo. Não é contrário à razão que eu escolha minha total destruição só para evitar o menor desconforto a um índio ou de uma pessoa que me é inteiramente desconhecida."[3]

Que filósofos competentes discordem tão profundamente sobre um assunto tão importante e tão discutido como este pode, certamente, ser explicado apenas por uma única hipótese: a saber, estão falhando em distinguir adequadamente entre os vários sentidos do termo "razão" e seus derivados. A presente discussão estabelece para si o modesto objetivo de remover ao menos um obstáculo no caminho do acordo, apontando uma distinção básica entre o significado do termo "racional" e o do termo "razoável". A maioria dos filósofos morais assume que esses dois termos (ou seus antônimos) são sinônimos em todos os contextos. Desejo salientar, entretanto, que, pelo menos, em alguns contextos, o termo "razoável" é usado com implicações que não se estendem ao termo "racional" e que, portanto, é necessário traçar uma distinção entre ambos os termos. Algumas das consequências desta distinção serão brevemente desenvolvidas.

2. A conduta irrazoável é necessariamente irracional? Vamos considerar a seguinte situação: dois indivíduos, *A* e *B,* têm uma pretensão igualmente forte por receber uma determinada quantia em dinheiro (talvez, por exemplo, uma comissão sobre a venda de certos bens, em cuja

[1] Texto de W.M. Sibley, originalmente publicado em *The Philosophical Review*, Vol. 62, No. 4 (January 1953) (Oct., 1953), pp. 554-560 (https://doi.org/10.2307/2182461). Agradeço à Duke University Press por autorizar a presente tradução. A distinção entre racional e razoável e razoável proposta por Sibley neste artigo é retomada por John Rawls ao sustentar que os cidadãos possuem duas faculdades morais, o racional e o razoável (cf. *O liberalismo político*, Conferência II, §1.1, nota 1.).

[2] Para uma excelente confirmação desta posição, ver a Part II de J. B. Pratt's Reason in the Art of Living (New York, Macmillan, 1949).

[3] A Treatise of Human Nature, II, iii, 3. N.T.: Utilizo aqui a tradução de Débora Danowski para a passagem de Hume citada por Sibley. Cf. HUME, David. Tratado da natureza humana. Tradução de Débora Danowski. 2. ed. São Paulo: UNESP, 2009, p. 452.

Recebido em 4 de setembro de 2025. Aceito em 29 de setembro de 2025.

venda ambos desempenharam um papel). O indivíduo *A*, no entanto, está em posição de reter todo o dinheiro para si mesmo e decide fazer isso sem levar em consideração os direitos de *B* em tal questão. Considerando que *A* esteja totalmente ciente do que está fazendo, como poderíamos caracterizar sua ação? Certamente, a caracterizaríamos como uma ação egoísta e, também, do ponto de vista moral, como errada. Assim, – especialmente se *B* (ou alguém que fale em seu nome) tivesse protestado contra a decisão de A, – nós atribuiríamos ainda à conduta de *A* um outro adjetivo: *irrazoável*. As consequências da ação de *A* sobre o bem-estar ou os direitos de *B* não constituem para *A* uma circunstância tal que ele esteja disposto a levar em conta na formação de sua decisão - exceto, talvez, na medida em que estas consequências possam afetar negativamente os próprios interesses de *A,* mesmo que de maneira indireta. O fato de *B* ser prejudicado não é considerado por *A* como uma razão suficiente para decidir agir de outra forma. Sendo egoísta, o indivíduo *A* não levará em consideração qualquer princípio a partir do qual o indivíduo *B* pudesse tentar argumentar com ele.

Entretanto, defendo que não necessariamente – ou, pelo menos, não imediatamente – chamaríamos *A* de irracional no sentido em que "irracional" significa "tolo", "absurdo", "ridículo", "sem sentido" ou "pouco inteligente". Supondo que o indivíduo *A* seja um egoísta empenhado em maximizar seu próprio bem-estar, estaríamos preparados para julgá-lo como tolo ou pouco inteligente apenas se ele fizesse uma estimativa incorreta dos resultados de seu próprio egoísmo e, ao fazê-lo, realmente prejudicasse o seu próprio bem-estar final. Nesse caso, seria correto acusar *A* de ter agido não apenas de maneira egoísta, mas também tola tanto do ponto de vista "irrazoável" quanto do "irracional". Que toda ação egoísta ou "injusta" seja também, em última análise, uma ação tola em termos de seu próprio bem-estar real é uma posição tão antiga quanto a *República.* E pode ser que seja uma posição sensata. Todavia, nosso ponto consiste no fato de que é necessário apresentar algum argumento adicional (por exemplo, um argumento baseado em uma análise da verdadeira natureza e das necessidades do ser humano) para sustentar essa posição – isto é, para mostrar que há uma conexão entre a mera insensatez intelectual e a atitude irrazoável demonstrada pelo homem egoísta. Em suma, condenar *A* como irrazoável não é *ipso facto* marcá-lo como irracional; e, portanto, esses dois termos não são, pelo menos neste contexto, sinônimos.

3. O significado de "racional". Antes de prosseguirmos, tentaremos especificar de maneira mais precisa os significados apropriados desses dois termos. Sugiro que o termo "racional", quando aplicado à conduta, seja utilizado com as seguintes implicações:

(A) (i) No que diz respeito aos *fins* que proponho para mim mesmo, "racional" implica: (a) que eu deveria ter uma compreensão informada da natureza dos fins que pretendo alcançar, incluindo nessa reflexão a compreensão do seu alcance na medida em que afetam outros fins – não apenas os meus, mas também os de outras pessoas afetadas por minhas ações; e (b) que, no caso de um conflito entre dois dos meus fins propostos, eu escolha aquele que realmente prefiro, ou seja, aquele fim que, após reflexão informada e cuidadosa, levando em consideração minha própria experiência e o que sei da experiência de outros, julgo ser de maior valor para mim.

(ii) No que diz respeito aos meios propostos para alcançar esses fins escolhidos racionalmente, implica que eu selecione aqueles meios que, com base nas melhores evidências disponíveis, são a forma mais eficaz para realizar esses fins; e que tomo conhecimento de todas as outras medidas que estão a meu alcance e que são necessárias para salvaguardar a realização de meus

fins.

(B) No que diz respeito à minha vontade, implica que eu ajo de acordo com as decisões tomadas por esse processo de reflexão, não permitindo que quaisquer influências emocionais me persuadam a seguir num caminho contrário.

Falhar em qualquer um ou em todos esses aspectos é ser irracional, no sentido de ser tolo, absurdo, pouco inteligente. Assim, eu me comporto irracionalmente quando não me preocupo em compreender a verdadeira natureza dos fins que estabeleço para mim mesmo; ou quando eu imprudentemente sacrifico um fim por uma segunda opção que, quando alcançada percebo como sendo de menor valor para mim do que a primeira teria sido; ou, quando seleciono meios irrealistas; ou, quando, tendo chegado a uma decisão suficientemente racional, falho em implementá-la na prática. Racionalidade – pelo menos nesses sentidos da palavra – é essencialmente uma virtude *intelectual*, embora também inclua secundariamente uma referência à vontade.

É pertinente observar que a mera caracterização de uma pessoa como racional não acarreta imediatamente qualquer informação a respeito de outras disposições ou fins dessa pessoa. As disposições egoístas ou altruístas, por exemplo, não são *per se* nem racionais nem irracionais. Tais adjetivos tornam-se aplicáveis às disposições ou ações apenas quando estas são vistas em relação a algum fim tomado pelo agente como último. Assim, pode ser irracional da minha parte preferir a destruição do mundo inteiro a arranhar meu dedo – mas apenas se, por exemplo, eu fosse o tipo de pessoa que realmente não deseja obter uma quantidade insignificante de bem-estar pessoal ao custo de calamidade para os outros; ou se, embora sendo um egoísta completo, não percebesse que a destruição do mundo inteiro poderia muito bem ter consequências piores para mim do que arranhar meu dedo. Não podemos caracterizar qualquer ação como racional ou irracional, a menos que possamos presumir as disposições ou propósitos que orientam o agente. Não é irracional de minha parte colocar meu braço no fogo – se meu objetivo é mutilar ou destruir a mim mesmo.

É evidente que, para agir racionalmente, devo levar em consideração todos os fatores relevantes; e dentre estes estará a reflexão sobre como aquilo que pretendo fazer irá repercutir sobre a satisfação dos interesses das outras pessoas. Caso contrário, dificilmente posso dizer que tenho uma compreensão inteligente do que estou fazendo. Há, contudo, uma diferença óbvia entre (1) levar em conta os interesses das outras pessoas tomando-as meramente como elementos na situação capazes de afetar a promoção de "meus próprios" interesses (onde os "meus próprios" interesses são opostos ou, pelo menos, distintos, dos interesses dos outros); e (2) levar em consideração os interesses dos outros como um espectador desinteressado e imparcial poderia fazer, ou seja, colocando-os em pé de igualdade com "meus próprios" interesses. Qualquer egoísta prudente leva em consideração os interesses dos outros no primeiro sentido. Mas levar em conta seus interesses no segundo sentido requer algo a mais do que simplesmente possuir um intelecto capaz de calcular corretamente as consequências futuras. Exige uma disposição genuinamente empática *(sympathetic disposition)* para com as outras pessoas, uma prontidão em se preocupar sinceramente com os interesses "delas" em si mesmos, assim como com os meus, e uma disposição para ser "objetivo" não apenas em um sentido lógico, mas também em um sentido distintivamente moral. Se eu possuir essa virtude *moral,* não serei, então, meramente *racional,* mas também estarei disposto a agir, quando os interesses dos outros estiverem envolvidos, de maneira *razoável*.

4. O significado de "razoável". Quando julgamos que alguém agiu de maneira razoável, podemos ter em mente uma situação moral ou não moral. Por exemplo, quando dizemos que "O investimento de *C* não deu certo", poderíamos dizer, "mas o risco envolvido era razoável, e ele tomou todas as precauções razoáveis." Aqui, "razoável" significa, até onde posso ver, a mesma coisa que "racional": *C* agiu apenas após uma análise inteligente da situação, aceitou somente os riscos que, com base nas evidências, uma pessoa racional estaria disposta a aceitar e tomou as precauções que a prudência normalmente ditaria. Mas, em uma situação na qual os juízos morais são pertinentes, se eu desejo que minha conduta seja considerada *razoável* por alguém que toma o ponto de vista moral devo demonstrar algo mais do que mera racionalidade ou inteligência. Ser razoável, nesse contexto, é ver a questão – como costumamos dizer – do ponto de vista da outra pessoa, descobrir como cada uma seria afetada pelas possíveis ações alternativas; e, além disso, não apenas para "perceber" isso (pois qualquer pessoa meramente prudente o faria), mas também para estar disposto a ser desinteressadamente *influenciado* ao tomar uma decisão a partir da avaliação desses possíveis resultados. Devo justificar minha conduta com base em um princípio ao qual todas as partes interessadas possam apelar e a partir do qual possamos raciocinar em conjunto. Se busco, por exemplo, justificar minha ação apontando algum bem que ela produz para mim, devo estar preparado para aceitar, ao menos como uma objeção *prima facie*, que ela tem como resultado algo que o outro julga prejudicial a si mesmo. A razoabilidade, portanto, requer imparcialidade, "objetividade"; expressa-se na noção de equidade *(equity)*. Essa exigência é, acredito, a essência do princípio de universalidade de Kant. A alternativa a ela é apenas o recurso à força.

Embora a fórmula kantiana possa expressar a essência da "razoabilidade", é inútil, contudo, tentar, como fez Kant, derivar a noção de comportamento razoável da noção de mera racionalidade. Neste sentido, somente agirei razoavelmente se eu tiver o *desejo* de ser razoável. A razão pode servir a essa "paixão" como serve a outras paixões, indicando para mim o que é exigido em minha conduta, caso esse seja o fim que me propus a alcançar. Mas a razão não escolhe esse fim, e não pode me oferecer quaisquer razões para ser razoável. A razão só entra em jogo quando algum fim já foi proposto. Certamente, se pretendo ser um egoísta, a razão pode aconselhar prudência em meu egoísmo; posso ser advertido a me comportar – ao menos externamente – de maneira razoável por medo das penalidades que enfrentaria se não o fizesse. No entanto, o que a razão me diz, então, não é simplesmente: "Seja razoável!" mas sim: "Seja razoável – se for necessário!" Ela emite apenas imperativos hipotéticos.

Hume está, portanto, correto ao ver que a moralidade não decorre unicamente da razão. A moralidade surge de uma disposição distinta, que pode ou não coexistir com a inteligência. Essa disposição não pode ser inculcada na mente de quem a não tem apenas por meio de argumentos, embora possa ser incutida por outros métodos. Por outro lado, porém, é um erro supor que, quando buscamos uma justificativa moral para nossa conduta, abandonamos completamente o raciocínio e recorremos apenas a diversas técnicas de persuasão não racional. Uma vez que uma pessoa decide que deseja agir moralmente, a reflexão pode revelar com quais princípios ela está implicitamente se comprometendo e, assim, indicar-lhe quais proposições são relevantes – e quais são relevantes e suficientes – para o processo de justificação moral de sua conduta. Tendo desejado ser razoável, ela está, então, obrigada a apresentar razões – não emoções.

5. A conduta racional "maximiza valores"? Há um aspecto do assunto em questão que exige

uma discussão mais aprofundada. Afirma-se que uma pessoa racional "prefere o valor maior ao menor", e, a partir disso alguns escritores inferem que uma pessoa racional "maximizará valores" segundo a maneira proposta pela teoria utilitarista. Admito a afirmação, mas rejeito a inferência. Suponha – para modificar nosso exemplo anterior – que *A* está em uma situação de prosperidade e abundância, enquanto *B* se encontra em uma situação desesperadora, com pesadas responsabilidades; mesmo assim, tal como antes, *A* toma para si o dinheiro, em prejuízo de *B*. Não se poderia, então, propor o seguinte argumento: "Sem dúvida, *A* preferiu aqui o menor valor ao maior; não teria sido *melhor* se ele tivesse dado o dinheiro a *B*? E não é isso, o fracasso em conseguir o melhor, em maximizar os valores, a própria essência da irracionalidade?"

Contudo, expressões tão triviais como "preferir o valor maior ao menor", "maximizar valores" etc., devem ser usadas com cautela, ou resultarão em confusões notáveis. Essas expressões (pressuponho aqui) são sempre elípticas. Ou seja, devemos indagar *de qual ponto de vista* é melhor que *B* receba o dinheiro. Certamente, suponho, não seria melhor do ponto de vista do indivíduo egoísta *A*. Estamos de acordo que a ação de *A* implica a ruína de *B*; mas o fato de *que a perda do dinheiro significa a ruína de B* não significa nada para *A*. A *felicidade* de *B* não é, para *A*, de forma alguma, um valor positivo. A *riqueza* de *A* é a única coisa que *A* valoriza; do *seu* ponto de vista, os valores *foram* maximizados e não podemos dizer (pelo menos sem mais evidências) que ele tenha sido irracional.

Agora, concordo que emitiríamos um juízo como: "É melhor que *B* fique com o dinheiro." Mas quem somos "nós"? Para que tal juízo tenha qualquer sentido objetivo, somos implicitamente obrigados a assumir algum ponto de vista que seja *critério*. Localizamos esse critério no ponto de vista de um espectador *C* informado, imparcial e empático. Suas preferências tornam-se, para o juízo moral, as normativas ou padrão. É *C* quem prefere, e julga melhor, que *B* seja resgatado do infortúnio, ao invés de que o saldo bancário já expressivo de *A* seja aumentado ainda mais. Mas *C*, ao acusar *A* de deixar de fazer o que *C* considera melhor, não irá automaticamente julgar *A* como irracional. Ele julgará que *A* agiu erroneamente, e ele também pode, com propriedade, acusar *A* de não ser razoável, pois ser razoável é equivalente a estar disposto a resolver disputas como *C* as resolveria. Antes que *C* possa afirmar que *A* é irracional, entretanto, ele deve saber quais são as disposições que orientam a conduta de *A*. Se, como em nosso exemplo, as disposições de *A* são puramente egoístas, então *C* teria que demonstrar que *A* foi míope em seu egoísmo. Alternativamente, se *A* – para mudar um pouco o exemplo – possui um interesse genuíno em agir de maneira razoável, mas, por algum motivo falhou em fazê-lo nesta ocasião e mais tarde demonstra arrependimento, pode-se novamente dizer que ele agiu de forma irracional *(foolishly)*; pois o arrependimento é um sinal de que ele não agiu de acordo com suas preferências fundamentais.

Uma pessoa que é racional, então, não é *ipso facto* utilitarista. Enquanto racional, ela agirá de modo a alcançar aquilo que, para ela, representa um valor maior; fará aquilo que realmente prefere fazer. Mas esse fato não esclarece *o quê* ela realmente prefere fazer. Se, no entanto, ela preferir agir forma razoável, então, necessariamente, ao raciocinar sobre sua conduta, dará atenção ao sentido da máxima utilitarista. Enquanto o egoísta nunca ignora quem experimenta as "dores" e "prazeres" produzidos por uma ação, a pessoa razoável o faz, e, portanto, – as outras circunstâncias permanecendo iguais – o fato de que a ação *X*, por exemplo, causa uma "pequena dor" para *A*, mas um" grande prazer" para *B*, o influenciaria a aprovar *X*. O cálculo utilitarista talvez não

seja suficiente, mas certamente é pertinente.

6. Conclusão. As seguintes conclusões emergem de nossa discussão: (1) Saber que uma pessoa é racional não nos permite saber quais os fins que ela buscará em sua conduta; sabemos somente que, quaisquer que sejam esses fins, ela buscará realizá-los de maneira inteligente. (2) Saber, entretanto, que uma pessoa está disposta a agir razoavelmente no que diz respeito aos outros, nos permite inferir que ela está disposta a orientar sua conduta por um princípio de equidade (*equity*) a partir do qual ela e os demais podem raciocinar em comum; e também que ela, ao tomar suas decisões, aceitará como relevantes *per se* as informações relativas às consequências de suas ações sobre o bem-estar de outrem. A disposição para ser razoável não deriva e nem se opõe à disposição de ser racional. No entanto, é incompatível com o egoísmo, pois está essencialmente relacionada à disposição de agir moralmente.

W.M. Sibley

Universidade de Manitoba