

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 Bark¹
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 establish 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 Kant³. 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?

²Direct 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.

³For 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 intentions⁴.

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 benefits⁵.

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 establishing 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

⁶In 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 establishing of the State itself – *exeundum 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 establish 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 violations⁷.

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*. Berlin: 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.