

Os construtivismos kantianos e a construção dos princípios de justiça na obra de Rawls

Antonio Saturnino Braga

antoniofsbraga@uol.com.br

Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brasil

resumo O objetivo do presente artigo é defender as seguintes teses: em primeiro lugar, as teorias da justiça de Habermas e Rawls devem ser vistas como duas versões do construtivismo kantiano, entendido como abordagem metaética que pretende reunir antirrealismo e cognitivismo forte; em segundo lugar, o contraste com o “construtivismo reconstrutivista” de Habermas ajuda a esclarecer de que modo a teoria de Rawls atende à pretensão de universalidade própria do cognitivismo forte. Ao realçar o fato de que o construtivismo de Rawls está baseado no trabalho reflexivo da consciência individual, o contraste com Habermas permite compreender em que sentido os conceitos de equilíbrio reflexivo e consenso sobreposto designam um ideal com pretensão de validade universal, o que desmente o particularismo implicado nas interpretações coerentistas e relativistas destes conceitos. **palavras-chave** Justificação Moral; Construtivismo Kantiano; Habermas; Rawls; Posição Original; Equilíbrio Reflexivo

Entendido como posição no domínio da metaética, o construtivismo kantiano caracteriza-se pela ambição de reunir as três seguintes teses: antirrealismo em relação a propriedades e fatos especificamente morais, cognitivismo forte em relação à justificação de enunciados morais substantivamente determinados, e internalismo em relação à conexão entre razões epistêmicas (razões para crer na validade de determinados enunciados morais) e razões para agir, ou motivos (O’NEILL, 2003, p. 347-349). Este projeto está centrado no conceito kantiano de razão prática. Isto não implica que não se possa discutir até que ponto a filosofia moral do

Recebido em 14 de dezembro de 2012. Aceito em 30 de janeiro de 2013.

dois pontos, Curitiba, São Carlos, vol. 10, n. 1, p.11-36, abril, 2013

próprio Kant seria construtivista. Talvez seja possível defender a hipótese de que a teoria construtivista pressupõe um enquadramento da racionalidade prática mais intersubjetivista do que aquele que se deve atribuir à época e às obras do próprio Kant.

Independentemente do debate sobre esta questão, dois fatos são indiscutíveis. Em primeiro lugar, toda apropriação do conceito kantiano de razão prática efetuada no enquadramento intersubjetivista típico de nossa época tem necessariamente de ser construtivista. Em segundo lugar, admitindo-se que este enquadramento caracteriza-se pela tese do pluralismo das concepções de boa vida, as morais kantianas de nossa época, na medida mesmo em que têm de ser construtivistas, têm de conceder primazia ao conceito de justiça, em relação aos valores mais diretamente associados à noção de boa vida. Em nossa época, as interpretações da concepção kantiana de razão prática são antes de tudo teorias da justiça – o que não implica que elas não possam ser também, secundariamente, teorias sobre o bem, ou sobre o que é bom na vida humana.

Pode-se afirmar que as interpretações contemporâneas da moral kantiana bifurcam-se em duas grandes tradições, a rawlsiana e a habermasiana – já se pode falar aqui de tradições, considerando-se o grande número de filósofos que exercem sua energia intelectual e criativa com base nos escritos de Rawls e Habermas. Ao afirmarmos isso, estamos querendo dizer que, ao serem enfocadas do ponto de vista da metaética, essas duas tradições devem ser vistas como expressões do construtivismo kantiano em geral. Mais precisamente, estamos querendo dizer que, desse ponto de vista, o “reconstrutivismo” de Habermas deve ser visto como uma forma de interpretar e realizar as ambições construtivistas acima mencionadas.

Admitindo-se que a abordagem construtivista pressupõe um enquadramento grosso modo intersubjetivista da racionalidade prática, é forçoso admitir também que este enquadramento permite diferentes concepções sobre a relação entre a consciência individual e o contexto intersubjetivo no qual se insere a racionalidade prática, e sobre o papel que a teoria moral pode e deve exercer neste relacionamento. O enquadramento intersubjetivista significa que qualquer exercício individual da racionalidade prática tem de levar em consideração demandas e interesses dos outros indivíduos, frequentemente conflitantes com as nossas próprias demandas e interesses, e tem, necessariamente, de orientar-se para um entendimento

ou acordo com os outros indivíduos. Mas há duas maneiras de conceber as condições ou exigências constitutivas deste exercício grosso modo intersubjetivista da racionalidade prática, ou do procedimento deliberativo no qual ele se encarna.

Em primeiro lugar, mesmo admitindo que tais condições apresentam um aspecto ideal ou idealizado, o teórico moral pode concebê-las como idealizações em certo sentido enraizadas nas interações comunicativas e sociais formadoras de cada consciência individual, cuja reflexão, por isso mesmo, se exerce necessária e universalmente a partir de, e por meio de, interações dialógicas concretas com os outros (McCARTHY, 1994, p. 45-49). Desse ponto de vista, o aspecto ideal das condições remete à dinâmica de aperfeiçoamento contínuo dos pressupostos grosso modo normativos das interações comunicativas formadoras da consciência individual, um aperfeiçoamento que diz respeito tanto ao modo de operacionalizar estes pressupostos quanto ao grau em que eles são efetivamente cumpridos. Por outro lado, numa segunda interpretação das condições ideais do procedimento racional, o teórico moral pode remetê-las primordialmente à consciência individual que, partindo de sua experiência das relações com os outros, reflete sobre o modo idealmente correto ou justo de considerar as demandas e pontos de vista dos outros e chegar a um acordo com os outros (SCANLON, 1998, p. 189-206). No primeiro caso, o solo em que germina e floresce o exercício intersubjetivista da racionalidade prática consiste em interações dialógicas reais e concretas, indefinidamente aprimoradas; no segundo caso, ele consiste na reflexão da consciência individual a respeito do modo maximamente razoável de ponderar os interesses e pontos de vista dos outros.

No primeiro caso, o papel da teoria moral restringe-se a reconstruir a dinâmica de aperfeiçoamento contínuo de interações comunicativas e sociais anteriores a toda teoria. Nesse caso, em outras palavras, trata-se de reconstruir o modo pelo qual tal processo é desencadeado e sustentado por idealizações emanadas de tais interações e facticamente influentes nas mesmas. No segundo caso, o papel da teoria moral é mais propositivo: trata-se de propor à consciência individual as condições ideais do exercício intersubjetivista de sua racionalidade prática, definido como um exercício que, por orientar-se para um acordo racional (razoável) ou justo com os outros, precisa considerar as demandas e interesses dos outros de um modo

perfeitamente razoável ou justo. Nesse caso, em outras palavras, o papel do teórico moral é propor à consciência individual um modelo ideal de exercício da razão prática, esperando que esse modelo seja reflexivamente chancelado por cada consciência individual. Ora, admitindo-se que as duas concepções representam duas manifestações do construtivismo kantiano em geral, a primeira equivale ao “construtivismo reconstrutivista” da tradição habermasiana, e a segunda ao “construtivismo construtivista” da tradição rawlsiana.

Nosso propósito no presente trabalho é utilizar o contraste entre esses dois tipos de construtivismo para tentar esclarecer de que modo e até que ponto a teoria de J. Rawls cumpre a ambição construtivista de reunir antirrealismo e cognitivismo forte. É comum afirmar-se que ao longo de sua obra Rawls foi progressivamente abandonando a pretensão de validade universal própria do cognitivismo forte, em favor de uma posição, grosso modo, relativista, centrada nas convicções compartilhadas em uma cultura política específica. Se isto for verdade, será forçoso admitir que Rawls não pode ser considerado um construtivista (kantiano) pleno. Onora O’Neill, por exemplo, parece comprometer-se com este veredicto, ao afirmar que “nas últimas versões de sua teoria da justiça, Rawls apresenta uma concepção de cidadania democrática no interior de uma sociedade limitada como a fonte e contexto do raciocínio sobre a justiça” (O’NEILL, 2003, p. 360). Ora, nosso propósito no presente trabalho é utilizar o contraste entre os construtivismos de Habermas e Rawls para questionar esta interpretação de O’Neill e propor uma interpretação da obra de Rawls mais simpática ao construtivismo kantiano em geral.

O objetivo da própria O’Neill é examinar a evolução das noções de “construtivismo” e “construtivismo kantiano” na obra de Rawls, contrastando-as com o construtivismo kantiano que segundo ela se pode e deve atribuir ao próprio Kant. A conclusão da autora pode ser resumida nas afirmações de que “há razões para pensar que a concepção de Kant da justificação ética é mais radicalmente construtiva do que aquela que Rawls propõe” (O’NEILL, 2003, p. 354), na medida em que “Kant, ao contrário de Rawls, compromete-se com uma explicação fortemente construtiva não apenas da justiça, e de modo mais abrangente da ética, mas, de modo mais radical, da própria razão prática” (Idem, p. 362-363). Para a autora, o construtivismo de Rawls limita-se à construção dos

princípios de justiça por meio da concepção de racionalidade prática modelada no procedimento hipotético da Posição Original. Em relação à justificação da própria Posição Original, entretanto, as duas concepções propostas por Rawls, a do equilíbrio reflexivo e a do consenso sobreposto, não podem ser consideradas construtivistas, na medida em que esvaziam a pretensão de validade universal essencial ao cognitivismo forte típico do construtivismo pleno. Para O'Neill, a concepção do equilíbrio reflexivo equivale a uma fundamentação meramente coerentista da Posição Original, enquanto a ideia do consenso sobreposto enfraquece ainda mais a justificação oferecida por Rawls, aproximando-a de uma posição claramente relativista.

Em contraste com Rawls, Kant oferece, segundo ela, uma fundamentação construtivista da própria razão prática; mais precisamente, Kant oferece uma fundamentação construtivista do princípio supremo da razão prática, o imperativo categórico. Embora nosso objetivo não seja avaliar a interpretação do próprio Kant exposta por O'Neill neste artigo específico, vale destacar que os argumentos por meio dos quais ela contrasta Kant com Rawls parecem indicar que ela não leva na devida conta uma distinção que é essencial para a compreensão não apenas da fundamentação estritamente kantiana da razão prática, mas também do projeto do construtivismo kantiano em geral – e esta afirmação é válida independentemente da posição que se queira adotar a respeito da hipótese de que o enquadramento kantiano da razão prática não é suficientemente intersubjetivista para poder ser considerado construtivista. Formulando a questão em termos kantianos, trata-se da distinção entre a fundamentação da fórmula da lei universal como fórmula de aplicação do imperativo categórico e, por outro lado, a fundamentação da efetividade motivacional do mesmo como norma incondicionalmente válida para os sujeitos racionais finitos.

No contexto da teoria construtivista e/ou kantiana, fundamentar a razão prática é uma tarefa que se desdobra em dois níveis intimamente relacionados, mas logicamente distintos. Em primeiro lugar, é preciso justificar a fórmula ou critério por meio do qual se exerce a razão prática, que é a fórmula da possibilidade de universalização, ou, nas versões mais atuais marcadas pela primazia da intersubjetividade, a fórmula da aceitabilidade (“concordabilidade”) universal, associada ao critério da imparcialidade.

Em segundo lugar, é preciso fundamentar a efetividade normativa da razão pura prática, mostrar que ela não é mera fantasia dos teóricos kantianos. Nesse segundo nível, trata-se de mostrar que a razão prática pura tem autoridade e efetividade, ou seja, poder de afetar e motivar as atitudes, escolhas e atos dos sujeitos racionais finitos. Na obra de Kant o primeiro nível da fundamentação é desenvolvido, por exemplo, nas duas primeiras seções da *Fundamentação da Metafísica dos Costumes*, onde Kant justifica a fórmula da possibilidade de universalização extraindo-a do simples conceito de um dever categórico ou incondicionado, pressuposto como dado. O segundo nível é exemplificado pelo argumento desenvolvido na terceira seção, depois substituído pelo argumento do “fato da razão”; trata-se aí de fundamentar a efetividade normativa deste conceito de dever categórico, mostrando onde se enraíza seu poder de afetar e motivar as atitudes, escolhas e ações humanas.

A necessidade de atender de uma forma “internalista” a esses dois níveis de fundamentação é uma característica dos construtivismos kantianos em geral, na medida em que estes se apropriam do conceito kantiano de razão prática. Ora, no construtivismo especificamente rawlsiano, é preciso admitir que a concepção da Posição Original não atende a nenhum desses dois níveis de fundamentação. Ela atende a um terceiro nível, o nível da justificação de princípios de justiça conteudisticamente determinados. Em outras palavras, a Posição Original representa um modelo ideal de exercício da razão prática; enquanto tal, ela decerto representa uma estratégia de justificação dos princípios conteudísticos de justiça propostos por Rawls, mas não uma estratégia de fundamentação deste exercício – nem fundamentação do critério que o governa (o critério da concordabilidade universal, associado à imparcialidade), nem fundamentação da sua normatividade, ou seja, da sua autoridade e efetividade em relação às escolhas e ações humanas.

Para poder ser considerado um construtivista pleno, Rawls precisa cumprir a exigência de justificação da própria Posição Original. Em outras palavras, Rawls precisa cumprir a exigência de fundamentação do exercício da racionalidade prática idealizado no experimento hipotético da Posição Original, nos dois níveis em que essa exigência se desdobra. Ora, admitindo-se com O’Neill que os conceitos do equilíbrio reflexivo e do consenso sobreposto equivalem justamente a estratégias

de fundamentação da Posição Original, uma interpretação construtivista de Rawls precisa atacar a afirmação de que tais conceitos são, respectivamente, coerentista e relativista, ou seja, incompatíveis com a pretensão cognitivista forte típica dos construtivismos kantianos. Trata-se de mostrar que esses conceitos, assim como a relação entre eles, são compatíveis com a pretensão de validade universal própria do cognitivismo forte. É o que tentaremos fazer na terceira seção do presente trabalho, depois de proceder a uma análise um pouco mais detalhada do contraste entre os construtivismos de Habermas e Rawls.

Nosso propósito, como dito acima, é utilizar este contraste para defender uma interpretação plenamente construtivista de Rawls. A estrutura de nossa argumentação será constituída pelas seguintes hipóteses: o contraste com Habermas torna mais claro o fato de que o construtivismo de Rawls está centrado na consciência individual, no trabalho reflexivo da consciência individual. Este fato não anula a perspectiva intersubjetivista e o caráter grosso modo pragmático da justificação dos princípios de justiça, essenciais para todas as formas de construtivismo kantiano; por outro lado, o reconhecimento deste fato permite compreender que os conceitos de equilíbrio reflexivo e consenso sobreposto não se referem a um acordo essencialmente histórico e cultural, e nesse sentido particular, mas designam um ideal com pretensão de validade universal, concebido em termos de perfectibilidade indefinida do trabalho reflexivo da consciência individual.

II. Razão prática, construtivismo reconstrutivista e construtivismo construtivista

As abordagens construtivistas sustentam que, embora as prescrições e princípios morais possam ser justificados, e justificados num sentido forte ou exigente, tal justificação se desenvolve numa perspectiva antirrealista, rejeitando, portanto, o pressuposto de que há propriedades morais objetivas, no sentido de independentes dos sujeitos que usam o vocabulário moral. Assim, a primeira ambição do construtivismo moral é defender a possibilidade e viabilidade de um projeto de justificação que exclui o ideal de adequação a propriedades e fatos de uma realidade independente.

Mais precisamente, trata-se de defender a viabilidade de um projeto de justificação que exclui qualquer referência a um mundo objetivo, no sentido de independente dos sujeitos do conhecimento moral. Como veremos logo a seguir, tal projeto sem dúvida depende de um fator justificador que não está ao inteiro dispor dos sujeitos empíricos, e que nesse sentido é objetivo, mas tal fator, ao contrário do que ocorre com a justificação efetuada no âmbito do conhecimento do mundo natural, não remete a propriedades de uma realidade independente, mas consiste na própria razão prática pura, tomada como um poder simultaneamente interno aos sujeitos empíricos e superior às razões que lhes são simplesmente dadas, que podem ser chamadas de razões meramente subjetivas. Trata-se, por exemplo, de desejos, atitudes e crenças empiricamente dadas nas diferentes culturas.

Para esclarecer a ambição construtivista de combinar antirrealismo e cognitivismo forte, pode-se recorrer à teoria discursiva desenvolvida por Habermas. Ao propormos isso, estamos sugerindo que, numa interpretação mais flexível da forma discursiva de justificação prática, a teoria construtivista de Rawls pode ser considerada uma teoria discursiva. Nessa interpretação mais flexível, o discurso prático não necessariamente equivale a processos reais ou concretos de discussão com os outros, mas pode ser entendido como um processo de discussão hipoteticamente desenvolvido por um sujeito individual que idealmente dialoga com outros sujeitos, outros pontos de vista, tentando discernir princípios que esses outros sujeitos, tomados como pessoas perfeitamente racionais ou razoáveis, poderiam e deveriam aceitar. Nessa interpretação mais flexível, embora a reflexão do sujeito razoável exija abertura e experiência em relação às demandas, interesses e pontos de vista dos outros sujeitos em geral, ela não exige participação real em discussões concretas, como ocorre na interpretação estritamente habermasiana da forma discursiva de justificação. No terreno constituído pela experiência da relação com os outros em geral, o “discurso” rawlsiano constitui uma espécie de discussão ideal da consciência consigo mesma, uma discussão entre a consciência empiricamente condicionada e os apelos da razão pura.

Assim, para esclarecer a ambição construtivista de combinar antirrealismo e cognitivismo forte, recorreremos de início a elementos conceituais próprios da teoria discursiva desenvolvida por Habermas, destacando

logo a seguir as características pelas quais o “discurso” prático rawlsiano distingue-se do discurso estritamente habermasiano. Na etapa inicial, tentaremos esclarecer a ambição construtivista recorrendo às análises de Habermas sobre as diferenças entre duas formas de justificação discursiva, próprias, respectivamente, do âmbito do conhecimento teórico e do âmbito do conhecimento moral. Com efeito, a ambição construtivista, que é própria do discurso prático, fica mais clara ao ser contrastada com uma forma de justificação que, apesar de discursiva ou intersubjetivista, não é antirrealista, ao contrário, envolve uma concepção grosso modo realista a respeito das propriedades do mundo natural.

Com base em análises desenvolvidas por Habermas, pode-se afirmar o seguinte: uma teoria discursiva representa, simultaneamente, uma concepção da justificação de nossos enunciados e uma concepção da validade cognitiva gerada por essa justificação (HABERMAS, 1996, p.34-39). No âmbito da justificação dos nossos enunciados, a teoria discursiva equivale a uma concepção pragmática da justificação, que pretende substituir a concepção semântica. Para a concepção semântica, uma sentença é justificada quando pode ser logicamente derivada de sentenças básicas individualmente especificáveis, cuja verdade é por sua vez imediatamente acessível à percepção ou intuição. Para a concepção pragmática, em contrapartida, a justificação consiste numa prática essencialmente intersubjetiva, na qual o valor justificatório das intuições depende de razões intersubjetivamente partilhadas ou partilháveis, ou seja, razões de caráter essencialmente intersubjetivo.

No âmbito da teoria discursiva, a concepção pragmática da justificação associa-se à concepção epistêmica da validade cognitiva de nossos enunciados, que pretende substituir concepções grosso modo correspondentistas, entre as quais um habermasiano poderia incluir a concepção defendida por E. Tugendhat em *Propedêutica Lógico-Semântica* (TUGENDHAT, 1983, cap. 13). Embora admita que os fatos são entidades linguisticamente mediadas, ou seja, mediadas pela aplicação correta de predicados linguísticos, a concepção de Tugendhat baseia os fatos nas percepções de indivíduos linguisticamente competentes, ou seja, competentes na aplicação de predicados elementares. Ora, um habermasiano poderia argumentar que, ao basear os fatos em percepções individuais, independentes das discussões intersubjetivas, este tipo de concepção acaba por reintroduzir, ainda que

de forma indireta e, nesse sentido, disfarçada, a ideia da correspondência à realidade independente – independente, a saber, das divergências, discussões e acordos dos sujeitos do conhecimento. A concepção epistêmica, em contrapartida, além de reforçar a tese de que fatos equivalem a enunciados cognitivamente válidos, pretende depurar a noção de validade cognitiva de todo sentido correspondentista, apresentando-a em termos de aceitabilidade racional e interpretando pragmaticamente a noção de aceitabilidade racional, ou seja, situando tal noção no quadro eminentemente intersubjetivo da discussão ou discurso.

Para dar continuidade ao contraste entre as duas espécies de justificação discursiva, próprias, respectivamente, do âmbito teórico e do âmbito prático, vale a pena complementar as análises mais recentes de Habermas com uma recuperação da noção de “interesses de conhecimento”, introduzida por ele no final da década de sessenta. Usando esta noção de uma forma relativamente livre, independente do modelo de teoria crítica tentado por Habermas antes da sua “virada linguística”, pode-se afirmar o seguinte: situar a aceitabilidade racional no quadro essencialmente intersubjetivo da discussão implica afirmar que sentido e critérios da aceitabilidade racional têm a ver com os interesses humanos (intersubjetivamente partilhados) subjacentes aos diferentes âmbitos do discurso. No âmbito do discurso teórico, o interesse fundamental é o interesse de êxito ou sucesso na lida com o “mundo objetivo”, entendido como realidade que não está ao nosso inteiro dispor. Ora, este tipo de interesse acarreta pressupostos, grosso modo, realistas, que estão totalmente ausentes do âmbito do conhecimento prático ou moral. Neste último, com efeito, o interesse fundamental pode ser formulado como interesse na *construção* de um mundo social justo, ou seja, interesse na reorganização do mundo social segundo princípios de justiça *construídos* pela razão prática intersubjetivamente exercida.

Para Habermas, dizer que os princípios de justiça corretos são *construídos* pela razão prática intersubjetivamente exercida equivale a dizer que eles são *descobertos* em discernimentos morais próprios desta nossa faculdade (HABERMAS, 1996, p. 38-39). Entendendo a noção de construção em termos de “aquilo que está ao nosso dispor”, pode-se afirmar que cabe a nós construir um mundo social justo, ou reorganizar o mundo social segundo princípios de justiça corretos. Entendendo a noção de descoberta

em termos de “aquilo que não está ao nosso dispor”, pode-se afirmar que o discernimento dos princípios de justiça corretos é uma questão de descoberta, na medida em que provém de condições racionais que em certa medida se impõem a nós – o que significa, justamente, que elas não estão ao nosso inteiro dispor. Tais condições são expressões da razão prática pura intersubjetivamente orientada, tomada como faculdade simultaneamente interna aos sujeitos empíricos e superior às convicções, preferências e objetivos que lhes são simplesmente dados. O aspecto de descoberta equivale ao sentido estritamente cognitivo ou epistêmico dos enunciados de justiça corretos (lembrando que, no vocabulário de Habermas, a correção representa uma forma de validade cognitiva análoga à verdade. No vocabulário de Rawls, interpretado em termos de cognitivismo forte, o adjetivo correspondente seria “razoável”).

Entretanto, na medida em que a razão prática pura é em certo sentido interna aos sujeitos do conhecimento moral, a descoberta efetuada nos discernimentos puramente racionais pode ser considerada uma espécie de construção, efetuada pela racionalidade intersubjetivamente orientada ou exercida – desde que se depure o conceito de construção da ideia de “aquilo que está ao nosso dispor”. Esta espécie de interpenetração entre construção e descoberta é expressa por Habermas na seguinte afirmação: “O sentido *construtivista* dos juízos morais, compreendido segundo o modelo da autolegislação, não deve ser esquecido; mas tampouco se deve apagar o sentido epistêmico das justificações morais” (HABERMAS, 1996, p. 39. Grifo meu).

Apesar de baseada em noções habermasianas, a exposição geral do construtivismo feita até aqui pode ser considerada compatível com a tradição rawlsiana, ou, pelo menos, com a leitura estritamente cognitivista da obra de Rawls, que é a que estamos adotando neste trabalho. Como afirmado anteriormente, gostaríamos agora de passar para um segundo momento de nossa exposição, aquele no qual as duas tradições se separam.

No âmbito do construtivismo kantiano em geral, a tese de que os princípios morais corretos ou razoáveis são construídos pelos sujeitos desdobra-se em dois momentos complementares. Em primeiro lugar, normas justas da interação social são construídas por meio de um procedimento de discussão ou deliberação que envolve demandas e pontos de vista conflitantes, colocados por diferentes sujeitos. Em segundo lugar, a construção

destes princípios equivale ao acordo (entendimento) intersubjetivo alcançado neste procedimento grosso modo discursivo. Como já foi afirmado, este procedimento inclui um aspecto ideal, mas há duas interpretações divergentes desta idealidade. Na interpretação habermasiana, a idealidade consiste na perfectibilidade indefinida da operacionalização e cumprimento dos pressupostos normativos das discussões práticas em geral. Esta perfectibilidade equivale à dinâmica de desenvolvimento da razão prática dos sujeitos comunicativamente relacionados, centrada na possibilidade (real) da razão libertar-se das determinações, grosso modo, heterônomas, ou seja, deixar de ser empiricamente condicionada e tornar-se razão prática pura (HABERMAS, 1999, p. 305).

No construtivismo habermasiano, o procedimento construtivo é ideal no sentido de estar orientado para a plena realização de idealizações emanadas das discussões práticas concretas, e facticamente influentes nas mesmas. Este sentido da idealidade do procedimento não implica que os sujeitos que devem empreendê-lo sejam sujeitos idealizados, ou sujeitos que apenas hipoteticamente se elevam ao ponto de vista da razão prática pura. Ao contrário, na versão habermasiana o procedimento ideal é apresentado como um procedimento a ser concretamente realizado pelos sujeitos empíricos, e que será realizado à medida que estes cumprirem de modo integral ou perfeito certas condições ideais sugeridas pelas – e nesse sentido presentes nas – pressuposições normativas das discussões práticas em geral. A realização plena do procedimento ideal (ou a realização plena da razão prática pura) é por sua vez um ideal de que os sujeitos empíricos podem concretamente se aproximar. No construtivismo habermasiano, portanto, princípios corretos são enunciados cognitivamente válidos produzidos (discernidos e construídos) pelos sujeitos concretos em condições ideais. Não cabe ao teórico moral propor nenhuma hipótese a respeito de quais seriam estes princípios. Ele deve aguardar o veredito a ser produzido pela discussão dos sujeitos concretos.

Se as duas tradições do construtivismo kantiano admitem a tese de que a construção-discernimento das normas corretas implica algum tipo de idealização, na tradição rawlsiana tal idealização deixa de ser referida primordialmente às interações comunicativas que enquadram e enformam os sujeitos concretos, e passa a ser referida primordialmente aos sujeitos (individuais) que empreendem hipoteticamente o procedimento

deliberativo de orientação intersubjetiva. O procedimento ideal é um procedimento empreendido por sujeitos ideais, ou seja, sujeitos que encarnam o ideal da razão prática pura. No vocabulário mais específico de Rawls, trata-se dos participantes da Posição Original. Trata-se de sujeitos hipotéticos; mais precisamente, trata-se de uma posição à qual os sujeitos concretos hipoteticamente se transportam ao refletirem sobre o modo maximamente razoável ou justo de deliberar sobre princípios de justiça. Tanto os sujeitos concretos quanto os sujeitos ideais em que os concretos hipoteticamente se transformam partem do pressuposto de que deliberar sobre princípios de justiça implica considerar as demandas e pontos de vista dos outros sujeitos, de modo a alcançar um entendimento essencialmente intersubjetivo. A Posição Original representa um modelo do modo correto ou razoável de realizar este exercício de consideração das demandas e interesses dos outros, tomado como um exercício que, embora orientado para o entendimento intersubjetivo, é realizado por sujeitos essencialmente individuais, de um modo hipotético, no terreno do pensamento. A sede da razão prática pura é o sujeito individual, não a relação intersubjetiva – mas a racionalidade deste sujeito consiste na capacidade de dialogar hipoteticamente com outros sujeitos em geral, de alcançar um acordo hipotético com sujeitos que apresentam interesses e pontos de vista divergentes daqueles a que ele empiricamente está inclinado.

No construtivismo habermasiano a racionalidade prática é situada no plano das interações comunicativas que desde sempre enquadram e enformam os sujeitos concretos. O procedimento ideal equivale nesse caso à realização plena de exigências fortes em certo sentido presentes em tais interações comunicativas – no sentido, a saber, de condições ideais que os participantes em alguma medida reconhecem, e que por isso são facticamente influentes. No construtivismo rawlsiano, em contrapartida, a racionalidade prática é situada no plano dos sujeitos individuais que empreendem hipoteticamente o procedimento intersubjetivo de construção dos princípios de justiça.

A diferença entre o construtivismo *reconstrutivista* de Habermas e o construtivismo *construtivista* de Rawls é, portanto, uma diferença nas tarefas respectivamente assumidas em cada uma das tradições. Na tradição habermasiana o teórico moral assume uma função meramente reconstrutiva: trata-se apenas de reconstruir exigências fortes presentes e em

certa medida reconhecidas nos processos comunicativos e discursivos que prefiguram o procedimento intersubjetivo plenamente racional. Na tradição rawlsiana, em contrapartida, o teórico moral assume duas funções em certo sentido construtivas. Sua “construção” equivale à colocação de duas propostas para os sujeitos concretos. Em primeiro lugar, ele propõe aos sujeitos um modelo ideal de exercício da racionalidade prática. Em segundo lugar, uma vez que a plausibilidade deste modelo para os sujeitos concretos depende dos princípios de justiça que seriam construídos pelos sujeitos ideais que o empreendem, cabe ao teórico moral apresentar aos sujeitos concretos uma hipótese a respeito de quais princípios seriam acordados por estes sujeitos ideais.

Na teoria de Rawls, estas duas funções construtivas são realizadas por meio do experimento hipotético da Posição Original. Cabe enfatizar que a realização dessas funções construtivas é algo que o teórico faz ao colocar determinadas propostas para a apreciação reflexiva dos sujeitos concretos. O teórico não constrói os princípios de justiça; ele propõe aos sujeitos concretos um modo de construí-los. A construção realizada pelo teórico é a construção de uma proposta a ser reflexivamente endossada pelos sujeitos concretos. Isto significa que o construtivismo rawlsiano não se esgota no modelo construtivo representado na Posição Original; ele inclui também um método de justificação deste modelo para os sujeitos concretos. Pode-se afirmar que os conceitos de equilíbrio reflexivo e consenso sobreposto definem o método defendido por Rawls. A questão é: será que estes conceitos atendem de modo satisfatório à pretensão de validade universal típica do cognitivismo forte?

III. Construtivismo, Posição Original, Equilíbrio Reflexivo e Consenso Sobreposto

Ao propor um modelo de exercício da razão prática para apreciação reflexiva dos sujeitos concretos, Rawls se deixa orientar por duas preocupações. Em primeiro lugar, é preciso que o modelo possa ser justificado para estes sujeitos. Para Rawls, esta justificação não pode espelhar-se na ideia kantiana de uma derivação puramente analítica a partir do conceito de dever ou lei como forma geral da moralidade. Ela tem de recorrer

às crenças conteudísticas sobre a moralidade adotadas pelos sujeitos aos quais ela se dirige. Ela equivale a uma espécie de equilíbrio entre tais crenças e o modelo ideal proposto, um equilíbrio “reflexivo”. Como veremos logo a seguir, o equilíbrio reflexivo não implica que a justificação tenha de ficar presa ou subordinada a tais crenças; implica apenas que o trabalho de justificação é um trabalho que opera sobre (a partir de) o solo destas crenças. Mais precisamente, de acordo com a interpretação que será defendida mais à frente, na concepção do equilíbrio reflexivo a justificação aparece como um equilíbrio indefinidamente aprimorável. A ênfase da tarefa de justificação não recai tanto no convencimento dos sujeitos empíricos, mas na sensibilização dos mesmos para as exigências e apelos da razão prática pura.

Considerando-se que o modelo da Posição Original deve ser justificável para todos os sujeitos aos quais é proposto, compreende-se a primeira preocupação de Rawls: trata-se de propor um modelo o mais delgado possível, ou seja, que incorpore exigências e condições as mais simples e plausíveis possíveis, ou as mais fracas possíveis (RAWLS, 1971, p. 20). Por outro lado, porém, as condições incorporadas no modelo precisam ser fortes o suficiente para gerar um resultado bem determinado, ou seja, para fundamentar princípios capazes de efetivamente orientar a resolução das questões de justiça. Condições demasiado delgadas podem ser insuficientes para a fundamentação de princípios determinados e interessantes. Tomemos, por exemplo, as condições da aceitabilidade universal e da imparcialidade. Elas certamente estão incorporadas no modelo representado na Posição Original. De acordo com este modelo, com efeito, o exercício da razão prática orienta-se pelo critério de que princípios corretos são princípios que podem ser aceitos por todos os possíveis concernidos, ou seja, que podem ser objeto de um acordo ou contrato entre todos os concernidos. E os sujeitos (ideais) representados no modelo deliberam de forma absolutamente imparcial, eles ficam impedidos de avaliar os possíveis princípios segundo a conveniência ou inconveniência dos mesmos para seus interesses e posições particulares. No modelo da Posição Original, a condição da imparcialidade é incorporada sob a forma do “Véu de Ignorância” (RAWLS, 1971, p.18-19).

Para Rawls, as condições da aceitabilidade universal e da imparcialidade são suficientemente fracas para aparecerem como devidamente justificadas

para todos os sujeitos concretos aos quais ele se dirige (e aqui é preciso novamente antecipar uma tese a ser defendida mais à frente, a saber, de que a limitação do âmbito dos destinatários da teoria, com suas crenças específicas, não implica subordinação às crenças destes destinatários, mas significa apenas que o trabalho de justificação precisa operar a partir das crenças de algum conjunto de destinatários, consistindo numa espécie de “transcendência desde dentro” destas crenças culturalmente específicas). Mas a vantagem pelo lado da justificação diante dos destinatários equivale a uma desvantagem pelo lado da fundamentação de princípios determinados e interessantes. Para Rawls, tomadas por si sós, as condições da aceitabilidade universal e da imparcialidade não são suficientemente fortes para permitirem a identificação ou construção de princípios bem determinados. É por isso que, na Posição Original proposta em *Uma Teoria da Justiça*, elas são complementadas por dois outros critérios.

Rawls propõe, em primeiro lugar, um critério referido aos interesses que movem os sujeitos ideais (hipotéticos) que participam da deliberação modelada na Posição Original. De acordo com este critério, os participantes são movidos pelo interesse em ter a maior quantidade possível de “bens primários”, entendidos como meios necessários ou úteis para a persecução de quaisquer fins, ou de qualquer concepção de boa vida (RAWLS, 1971, p. 92-93). A partir desse critério, que se associa à característica do desinteresse mútuo dos sujeitos hipotéticos representados na Posição Original, a deliberação destes sujeitos assume a feição de uma escolha daqueles princípios capazes de assegurar-lhes o nível mais satisfatório possível de bens primários. Ora, ao ser configurada pelo Véu de Ignorância como operacionalização da exigência de imparcialidade, a escolha dos sujeitos hipotéticos assume a feição de “escolha racional em condições de incerteza” (RAWLS, 1971, p. 152-156). E disto resulta um segundo critério para complementar as condições da aceitabilidade universal e da imparcialidade: o critério de escolha *maximin* (em condições de incerteza quanto à probabilidade das diversas situações possíveis, a escolha racional é aquela que focaliza a pior situação possível e procura garantir o melhor resultado nessa situação).

Na obra de 1971, portanto, Rawls propõe o seguinte modelo de exercício da razão prática, orientado para o discernimento de princípios de justiça corretos. Princípios corretos são aqueles que são universalmente

aceitos ou acordados pelos participantes de uma deliberação justa, ou seja, uma deliberação que resolve de maneira imparcial os conflitos intersubjetivos de interesses, finalidades e pontos de vista. Os participantes (hipotéticos) da deliberação justa estão interessados apenas em garantir para si próprios a maior quantidade possível de bens primários, e não têm interesse ou invejada quantidade alocada aos demais participantes. Ao mesmo tempo, eles ignoram todas as características que, apesar de essenciais na determinação da situação e chances que respectivamente possuem na ordem social concreta, como riqueza socialmente herdada e talentos naturalmente concedidos, devem ser consideradas casuais e irrelevantes no que tange à determinação das normas de justiça. Esta ignorância implica que, apesar de quererem escolher os princípios mais favoráveis a si próprios, sua escolha se realiza em condições de absoluta incerteza. Por isso, os participantes da deliberação justa adotam o critério de escolha maximin. Adotando este critério, eles terminam por estabelecer um acordo em torno dos dois princípios da “justiça como equidade”, constitutivos de uma espécie de liberalismo igualitário.

Para o Rawls de 1971, esta proposta atende às duas preocupações por ele focalizadas. Pelo lado da fundamentação dos princípios de justiça, as condições e critérios que estruturam o modelo são suficientemente densos para permitirem o discernimento ou construção de princípios determinados e interessantes. Já pelo lado da justificação da proposta diante dos sujeitos concretos, tais condições e critérios são relativamente delgados ou fracos, eles não exigem dos sujeitos concretos concessões demasiado moralísticas, na medida em que configuram a deliberação justa em termos de escolha racional e auto interessada. E isto significa que o modelo por eles estruturado poderá justificar-se diante dos sujeitos concretos que o teórico procura sensibilizar. Se a tarefa de justificação é antes de tudo uma tarefa de sensibilização dos sujeitos empíricos para as exigências e apelos da razão prática pura, no modelo de 1971 tais apelos recaem, sobretudo, sobre o igualitarismo expresso no “princípio da diferença”, contra concepções de justiça menos igualitárias, baseadas numa priorização quer do mérito quer da eficiência social (SCANLON, 2003, p. 156-157). Para o Rawls de 1971, o igualitarismo da “Justiça como Equidade” ia além das convicções sobre a justiça dominantes entre os sujeitos aos quais ele se dirigia, e tratava-se de justificar este apelo moralmente mais denso e

exigente “equilibrando-o” com concepções grosso modo egoísticas, mais alinhadas às convicções dominantes, e incorporadas no modelo sob a forma da configuração da deliberação justa em termos de escolha racional e auto interessada.

Esta compreensão da tarefa de sensibilização vai se alterar. Para o Rawls de *Liberalismo Político* (1993 e 1996), o alvo primordial deixa de ser a sensibilização dos sujeitos para um igualitarismo frontalmente contrário a concepções meritocráticas e utilitaristas do valor do ser humano, e passa a ser a sensibilização para os deveres de tolerância, razoabilidade e civilidade em relação a pessoas que abraçam doutrinas sobre o bem contrárias à do sujeito individualmente tomado. Se o modelo de deliberação justa é um modelo de deliberação razoável, no último Rawls a noção de razoabilidade deixa de ser operacionalizada em termos de restrições e exigências impostas à escolha racional e autointeressada, e passa ser operacionalizada em termos de disponibilidade para abrir-se ao ponto de vista do outro, para compreender e até certo ponto anuir às razões do outro. Numa formulação alternativa, a razoabilidade passa a ser apresentada como a capacidade de identificar e utilizar um leque de razões comuns, capazes de promover a aproximação e entendimento entre partidários de doutrinas sobre o bem radicalmente divergentes.

Pode-se discutir até que ponto essa mudança não tornaria dispensável o conceito da Posição Original. Talvez seja possível afirmar que este conceito é logicamente dependente da modelagem da deliberação justa em termos de escolha racional e autointeressada, submetida a determinadas restrições e exigências. Mas não discutiremos esta questão aqui. Partiremos da interpretação mais ortodoxa de que em *Liberalismo Político* a Posição Original deve ser entendida como um modelo de deliberação entre pessoas perfeitamente (idealmente) razoáveis.

Analisemos agora o conceito do equilíbrio reflexivo. Ao longo de toda sua obra, Rawls mantém a tese de que a justificação da Posição Original diante dos sujeitos concretos realiza-se por meio de equilíbrio reflexivo (RAWLS, 1971, p. 17-22; RAWLS, 1980, p. 533-535; RAWLS, 1996, p. 22-28). Trata-se, mais precisamente, do equilíbrio reflexivo “amplo” e “geral” (RAWLS, 1996, p. 384, nota 16). Três questões são fundamentais para a compreensão das relações entre as noções de “equilíbrio reflexivo”, “equilíbrio reflexivo amplo” e “equilíbrio reflexivo geral”. Em primeiro

lugar, quais são os polos entre os quais se dá o equilíbrio? Em segundo lugar, qual o lugar ou domínio em que se dá o equilíbrio? Em terceiro lugar, em que consiste exatamente o equilíbrio?

Começemos com o equilíbrio reflexivo *tout court*. E começemos com a primeira questão. Com base na análise do equilíbrio reflexivo efetuada por T. Scanlon, pode-se afirmar o seguinte: partindo-se do princípio de que o equilíbrio reflexivo é um método de justificação do modelo proposto pelo teórico, e uma justificação dirigida, fundamentalmente, a cada sujeito concreto ao qual o modelo está sendo proposto, trata-se de um equilíbrio entre os seguintes polos (SCANLON, 2003, p. 139-157). Por um lado, o polo constituído pelas convicções sobre a justiça que em princípio gozam da aceitação e confiança do sujeito concreto ao qual o teórico se dirige. Trata-se de convicções em diferentes níveis de generalidade e âmbitos de aplicação. Podem ser convicções sobre ações e instituições específicas, ou convicções sobre normas mais gerais; pode tratar-se também de convicções sobre o modo justo de resolver conflitos de interesses e pontos de vista, ou seja, sobre as condições de uma deliberação justa (RAWLS, 1996, p. 25-26). O segundo polo consiste no modelo proposto pelo teórico: trata-se de um modelo de exercício da racionalidade prática, que inclui uma representação da deliberação justa e uma representação do resultado desta deliberação, ou seja, dos princípios de justiça que dela resultam, segundo o encaminhamento argumentativo defendido pelo teórico.

A diferença entre o equilíbrio reflexivo *tout court* e o equilíbrio reflexivo “amplo” refere-se à inclusão neste último de um terceiro polo: modelos de justiça alternativos ou concorrentes, ou seja, representações concorrentes da deliberação justa, ou dos princípios que mais plausivelmente resultariam da deliberação justa. Mas o lugar ou domínio em que se dá o equilíbrio reflexivo amplo é o mesmo em que se dá o equilíbrio reflexivo *tout court*: a consciência (individual) de cada sujeito concreto ao qual o teórico está se dirigindo. Ora, à medida que o equilíbrio reflexivo amplo alcançado em certa consciência individual é idêntico ao alcançado nas consciências de todos os demais sujeitos concretos aos quais o teórico está se dirigindo, ou seja, à medida que o domínio em que se dá o equilíbrio reflexivo deixa de ser a consciência isoladamente tomada e passa a ser o conjunto dos sujeitos concretos, o equilíbrio reflexivo “amplo” torna-se (também) “geral”.

O parêntese que se apôs ao “também” exprime uma tensão entre os equilíbrios reflexivos “amplo” e “geral”. Para ser geral o equilíbrio reflexivo talvez precise ser um pouco mais conservador do que aquele que pode ser alcançado na consciência individualmente tomada; talvez ele precise se prender um pouco mais às convicções sobre a justiça factualmente predominantes numa dada cultura.

Para compreender e resolver essa tensão, passemos à terceira das questões acima formuladas: em que consiste exatamente o “equilíbrio” do equilíbrio reflexivo? Uma resposta bastante comum corresponde à interpretação coerentista do equilíbrio reflexivo. Com efeito, ao entender a justificação propiciada por este método como uma questão de coerência na rede das crenças vigentes em dada comunidade linguística, a interpretação coerentista sugere que tal rede seja tomada como uma base em última instância dada e inultrapassável. E isto por sua vez sugere que as crenças e atitudes implicadas nessa rede sejam tomadas como pontos de apoio em última instância dados e fixos. Deste ponto de vista, o “equilíbrio” do equilíbrio reflexivo equivaleria a uma espécie de ajuste do modelo proposto pelo teórico às crenças empiricamente vigentes em dada cultura específica; a validade do modelo consistiria simplesmente no fato de que ele propicia maior organização, clareza e sistematicidade argumentativa à rede dessas crenças. Ora, se esta for a interpretação correta do equilíbrio reflexivo rawlsiano, será preciso concluir que Rawls abandona a pretensão de universalidade típica das teorias estritamente cognitivistas. Será preciso concluir, em outras palavras, que ele não pode ser considerado um construtivista pleno. Como vimos no início de nosso trabalho, esta é a opinião de Onora O’Neill.

Thomas Scanlon, entretanto, defende outra interpretação do equilíbrio reflexivo, mais simpática ao cognitivismo forte típico do construtivismo kantiano pleno (SCANLON, 2003, p. 149-153). A partir da interpretação proposta por Scanlon, pode-se afirmar o seguinte: as convicções inicialmente abraçadas pelo sujeito que aprecia a teoria, assim como a rede das mesmas, não devem ser tomadas como uma base em última instância dada, fixa e inultrapassável, mas como o solo sobre o qual opera o trabalho da reflexão racional, entendido como um movimento indefinidamente estendido de “ultrapassagem desde dentro” das crenças sobre a justiça adotadas pela consciência individual, em direção ao ideal de um encontro da

racionalidade historicamente vigente com a razão prática pura, a perfeita razoabilidade. Ora, este ideal, claramente, abriga a pretensão de validade universal. Mas não se trata de uma ideia já determinada, e nesse sentido já dada, e sim de um ideal a ser paulatinamente definido e redefinido, num processo de aprimoramento da descoberta daquilo que a razão pura exige de nós. Neste sentido, como já foi dito, a ênfase da tarefa de justificação não recai no convencimento dos sujeitos empíricos quanto a princípios definitivamente determinados, mas na sensibilização dos mesmos para as exigências e apelos da razão prática pura. O equilíbrio reflexivo é um equilíbrio indefinidamente aprimorável.

São úteis aqui duas citações do próprio Rawls.

Com relação ao modo como descobrimos o procedimento correto, o construtivista diz: por reflexão, usando os poderes da nossa razão. Mas como estamos usando nossa razão para descrever a si própria, e como a razão não é transparente a si própria, nós podemos descrever mal nossa razão, como qualquer outra coisa. O esforço para atingir o equilíbrio reflexivo continua indefinidamente (RAWLS, 1996, p. 96-97).

O equilíbrio reflexivo assemelha-se ao teste [de validade moral] de Habermas sob este aspecto: ele é um ponto no infinito que nós podemos nunca alcançar, embora possamos nos aproximar dele, no sentido de que por meio da discussão nossos ideais, princípios e juízos parecem-nos mais razoáveis, e nós os consideramos mais bem fundamentados do que eram antes (RAWLS, 1996, p. 385).

A menção à semelhança com Habermas não deve esconder o fato de que, no quadro da teoria rawlsiana, o fator essencial deste movimento contínuo para atingir a razoabilidade plena consiste na reflexão da consciência individual. O destinatário primeiro – ou árbitro último – do modelo proposto pelo teórico é a consciência individual (SCANLON, 2003, p. 152-153). Antes de promover uma discussão concreta da consciência individual com outros sujeitos concretos, trata-se para Rawls de promover a discussão ideal da consciência consigo mesma, uma discussão ou diálogo entre a consciência empiricamente condicionada e os apelos da razão pura. É claro que essas duas dimensões da discussão se interpenetram de forma praticamente inextricável, na medida em que a experiência da primeira é essencial à

segunda. Mas, a nota imediatamente anterior ao texto citado em último lugar deixa claro que, para Rawls, o equilíbrio reflexivo “amplo” (cujo domínio de realização é a consciência individual) é logicamente anterior ao equilíbrio reflexivo “geral” (cujo domínio de realização é o conjunto dos sujeitos concretos) (RAWLS, 1996, p. 384, nota 16).

Podemos retornar agora à tensão acima exposta entre os equilíbrios amplo e geral. Dizer que o destinatário primeiro do movimento reflexivo proposto por Rawls é a consciência individual, ou dizer que o equilíbrio reflexivo amplo é logicamente anterior ao geral – dizer isso não implica afirmar que Rawls não esteja interessado no acordo dos equilíbrios reflexivos individualmente alcançados. Lembremos que o equilíbrio reflexivo é um método de justificação de um modelo de acordo hipotético entre os sujeitos. Embora o acordo seja de início colocado no nível do modelo ideal ou hipotético, e não no nível dos sujeitos concretos que apreciam o modelo, a ideia do acordo reflete-se num interesse de que o equilíbrio reflexivo pelo qual cada indivíduo chancela o modelo ocorra de maneira mais ou menos idêntica em todos os demais indivíduos concretos.

É este interesse que explica os frequentes apelos ao caráter socialmente compartilhado das convicções sobre a justiça envolvidas no equilíbrio reflexivo. Pode-se afirmar que, quando o interesse no acordo dos equilíbrios reflexivos se acentua, Rawls correspondentemente enfatiza o caráter compartilhado das convicções sobre a justiça às quais ele apela. E isto pode de fato sugerir que tais convicções equivalham a uma base em última instância dada, um fundamento que se coloca além da discussão. Mas, a primazia teórica da consciência individual significa que, embora Rawls também esteja interessado no acordo dos equilíbrios reflexivos, seu interesse primário é o trabalho reflexivo da consciência individual. E isto, por sua vez, significa que as crenças socialmente compartilhadas, longe de constituírem um fundamento além da discussão, constituem o solo de um trabalho reflexivo que as engloba todas num movimento de superação dirigido ao ideal universal da razoabilidade plena.

A tensão entre os equilíbrios amplo e geral atinge em cheio o conceito do consenso sobreposto. É plausível afirmar que tal conceito designa um caso especial de equilíbrio reflexivo geral – trata-se do equilíbrio no qual coletivamente se endossa a modelagem da deliberação justa como deliberação entre sujeitos perfeitamente razoáveis, típica do último Rawls.

Ora, se o consenso sobreposto deve ser interpretado como um caso de equilíbrio reflexivo geral, a interpretação da tensão acima defendida implica que tal conceito deve ser entendido como um consenso logicamente subordinado à dinâmica reflexiva da consciência individual e à perfectibilidade contínua envolvida nesta dinâmica; subordinado, portanto, ao ideal de um encontro da racionalidade empiricamente condicionada com as exigências da razoabilidade plena. O consenso sobreposto não é um consenso concreto entre sujeitos empiricamente dados, baseado numa coincidência feliz (mas contingente) das convicções historicamente dadas numa cultura específica. Trata-se antes de um consenso entre sujeitos inseridos numa dinâmica reflexiva dirigida a um ideal, retratados como indivíduos que, apesar de compreenderem suas vidas e esforços pessoais segundo as valorações fortes típicas das doutrinas abrangentes do bem, e apesar de adotarem doutrinas diferentes e eventualmente conflitantes, compreendem-se simultaneamente como sujeitos dotados da capacidade de separar as questões sobre a boa vida das questões sobre a justiça. Em outras palavras, o consenso sobreposto é um consenso entre sujeitos atingidos pelos apelos de razoabilidade propostos pelo teórico, sujeitos que aceitam a tese de que, no âmbito das relações estritamente públicas ou políticas, considerações de justiça devem ter prioridade sobre as concepções do que é importante e valioso na vida pessoal (RAWLS, 1996, p. 388).

Num contexto em que o modelo a ser reflexivamente endossado consiste no ideal de uma deliberação entre sujeitos perfeitamente razoáveis, pode-se afirmar o seguinte a respeito dos equilíbrios reflexivos que são buscados e esperados: embora seja plausível esperar que o equilíbrio reflexivo “amplo” seja também “geral”, no sentido de concordante para todos os sujeitos concretos aos quais o trabalho reflexivo está sendo proposto, os caminhos para se chegar a este resultado serão admitidamente diferentes, em virtude das divergências nas convicções iniciais dos sujeitos que abraçam diferentes doutrinas abrangentes. E isto significa, como afirmado acima, que o equilíbrio reflexivo geral terá o caráter de um “consenso sobreposto” entre partidários (finitamente) razoáveis de diferentes doutrinas abrangentes. O consenso sobreposto exprime a concordância de resultado de diferentes caminhos de equilíbrio reflexivo (amplo).

Ora, admitindo-se que o consenso sobreposto equivale à concordância de resultado dos equilíbrios reflexivos amplos alcançados, alcançáveis e

indefinidamente aprimoráveis por pessoas (finitamente) razoáveis, ou seja, pessoas atingidas pelos apelos de razoabilidade propostos pelo teórico; admitindo-se, em segundo lugar, que este último tipo de equilíbrio reflexivo, à semelhança de todos os demais tipos, pode manter a pretensão de validade universal do ideal da razão prática pura (ainda que este ideal só exista na relação dialética com um polo oposto, constituído pelas convicções empíricas a serem reflexivamente trabalhadas e superadas); admitindo-se então estes pontos, pode-se concluir que os conceitos rawlsianos de equilíbrio reflexivo e consenso sobreposto são plenamente compatíveis com o cognitivismo forte típico do construtivismo kantiano.

Restaria analisar uma última questão. Até que ponto tais conceitos atendem à ambição internalista típica do construtivismo pleno? Até que ponto a teoria rawlsiana pode sustentar a tese de que os princípios justificados (ou justificáveis) no “consenso sobreposto” são intrinsecamente motivadores para sujeitos razoáveis *finitos*? Até que ponto a reflexão envolvida no ideal de um consenso sobreposto universal pode ser considerada intrinsecamente motivadora para os sujeitos finitos? Como fundamentar a autoridade motivacional dos princípios reflexivamente endossados, ou reflexivamente justificáveis? A análise dessas questões fica adiada para outra oportunidade.

Referências bibliográficas

_____. 1980. Kantian Constructivism in Moral Theory. *The Journal of Philosophy*, v. 77, n. 9, p. 515-572, September 1980.

_____. 1996. *Political Liberalism*. Paperback Edition. New York: Columbia University Press.

_____. 1998. *What We Owe to Each Other*. Cambridge, Massachusetts: Harvard University Press.

_____. 2003. Rawls on Justification. In: FREEMAN, S. (Ed.) *The Cambridge Companion to Rawls*. New York: Cambridge University Press.

- _____. 1999. *Verdade e Justificação*. Ensaios Filosóficos. Tradução de Milton Camargo Mota. São Paulo: Loyola, 2004.
- HABERMAS, J. 1996. *The Inclusion of the Other: Studies in Political Theory*. Tradução de Ciaran Cronin. Cambridge, Massachusetts: MIT Press, 1998.
- McCARTHY, T. 1994. Kantian Constructivism and Reconstructivism: Rawls and Habermas in Dialogue. *Ethics*, University of Chicago, n. 105, p. 44-63, October 1994.
- O'NEILL, O. 2003. Constructivism in Rawls and Kant. In: FREEMAN, S. (Ed.). *The Cambridge Companion to Rawls*. New York: Cambridge University Press.
- RAWLS, J. 1971. *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press.
- SCANLON, T. 1982. Contractualism and Utilitarianism. In: SEN, A.; WILLIAMS, B. (Eds.) *Utilitarianism and beyond*. Cambridge: Cambridge University Press.
- TUGENDHAT, E. 1983. *Propedêutica Lógico-Semântica*. Tradução de Fernando Rodrigues. Petrópolis: Vozes, 1996.

