

Contractualism and utilitarianism

Contratualismo e utilitarismo¹

T. M. SCANLON

(Autor)

MARCOS ANTONIO DA SILVA²

(Tradutor)

O utilitarismo ocupa um lugar central na filosofia moral de nossa época. Não é o ponto de vista que muitas pessoas detêm; certamente, há bem poucas pessoas que reivindicariam ser utilitaristas na ação. Mas para uma variedade muito ampla de pessoas, ele é o ponto de vista ao qual elas recorrem quando se acham pressionadas e quando tentam dar uma explicação teórica de suas crenças morais. Na filosofia moral, ele representa uma posição contra a qual alguém luta se se quer evitá-la. Isso é assim, a despeito do fato de que as implicações do utilitarismo na ação estão extremamente em desacordo com as convicções morais solidamente defendidas, embora o utilitarismo de regras, na sua formulação alternativa mais comum, atinja a maioria das pessoas como um compromisso instável.

O amplo apelo ao utilitarismo se deve, penso eu, às considerações filosóficas de um tipo, mais ou menos, sofisticado que nos impele a uma direção bastante diferente do que a nossa moral de primeira ordem acredita. Em particular, o utilitarismo deriva muito do seu apelo às alegadas dificuldades sobre os fundamentos dos pontos de vista rivais. O que uma alternativa de sucesso ao utilitarismo deve fazer, antes de tudo, é fragilizar essa fonte de energia, propiciando uma clara explicação sobre os fundamentos do raciocínio moral não-utilitarista.

¹ Sou muito grato a Derek Parfit pela crítica paciente e a discussão enormemente útil das muitas primeiras versões deste artigo. Agradeço também às inúmeras pessoas que ouviram algumas partes destas versões que foram proferidas em palestras e gentilmente respondidas com comentários pertinentes. Em particular, tenho uma dívida impagável com Marshall Cohen, Ronald Dworkin, Owen Fiss e Thomas Nagel pelas valorosas críticas.

² Graduado em Direito pela Universidade Estadual do Norte do Paraná - UENP - 2001; Especialização em Direito Internacional e econômico pela UEL - 2009. Mestre em Direito pela UENP - 2012. Doutorando em Filosofia pela Universidade Estadual de Londrina - UEL - 2021. ORCID: <https://orcid.org/0000-0003-0007-6630>. E-mail: marcosasilva789@gmail.com

Disso se segue que eu, em primeiro lugar, descreverei o problema detalhadamente, fazendo as perguntas que a explicação filosófica sobre os fundamentos da moralidade deve responder. Eu, então, apresentarei uma versão do contratualismo que, argumentei, oferece um conjunto melhor de respostas a essas perguntas mais do que aquele conjunto dado pelas versões assumidamente abertas do utilitarismo. Por fim, explicarei por que o contratualismo, como eu o entendo, não remete a nenhuma fórmula utilitária como sendo seu resultado normativo.

O contratualismo foi proposto como alternativa ao utilitarismo antes, sobretudo por John Rawls em *Uma teoria de justiça* (Rawls, 1971). Apesar da ampla discussão que esse livro recebeu, eu penso, no entanto, que o apelo ao contratualismo como ponto de vista fundacional foi mal avaliado. De um modo especial, não foi apreciado o suficiente o fato de que o contratualismo oferece uma explicação particularmente plausível sobre a motivação moral. A versão do contratualismo que eu apresentarei se distingue da de Rawls em vários aspectos. Em particular, ela não faz uso, ou apenas faz um tipo diferente ou mais limitado de uso, de sua noção de escolha por trás de um véu da ignorância. Um resultado dessa diferença é fazer com que o contraste entre contratualismo e utilitarismo fique mais claro.

362

I

Existe um tema, como o da filosofia moral, pela mesma razão que existem outros temas, como o da filosofia da matemática. Nos juízos morais, como nos juízos matemáticos, nós temos um conjunto de crenças putativamente objetivas, no qual somos inclinados a investir um certo grau de confiança e de importância. Refletindo, porém, não é nada óbvio sobre o que, se eles de fato existem, esses juízos podem ser e em virtude dos quais algumas coisas podem ser ditas para estarem certas ou serem defensáveis, enquanto outras não podem. Essa questão sobre o tema de estudo ou sobre os fundamentos da verdade, é a primeira questão filosófica sobre moralidade e matemática. Em segundo lugar, na moralidade e na matemática, parece ser possível descobrir a verdade simplesmente pensando ou raciocinando sobre ela. Experiência e observação podem ser úteis, mas a observação no sentido

normal não é o meio padrão de descoberta em nenhum dos objetos de estudo. Portanto, dada qualquer resposta positiva para a primeira questão - qualquer especificação do tema ou fundamento da verdade em matemática ou moralidade -, nós precisamos de alguma epistemologia compatível que explique como é possível descobrir os fatos sobre esse tema por meio de algo como os meios que nós parecemos usar.

Dada essa similitude em questões que dão origem à filosofia moral e à filosofia da matemática, não é surpresa que as respostas comumente dadas caiam em tipos gerais similares. Se nós fôssemos entrevistar estudantes recém-matriculados em um curso de matemática, a maioria deles, penso eu, se declararia favoráveis a algum tipo de convencionalismo. Eles defenderiam que a procedência da matemática vem de definições e de princípios que ou são arbitrários ou são instrumentalmente justificados, e que o raciocínio matemático consiste na percepção do que se segue dessas definições e desses princípios. Alguns outros, talvez, seriam realistas ou platonistas, de acordo com os quais as verdades matemáticas são um tipo especial de fato não-empírico que nós podemos perceber mediante alguma forma de intuição. Outros poderiam ser naturalistas que defendem que a matemática, adequadamente entendida, é apenas a ciência empírica mais abstrata. Por fim, há, embora talvez não entre estudantes recém-matriculados em um curso de matemática, os que defendem que não existe nenhum fato matemático no mundo “fora de nós”, mas que as verdades da matemática são verdades objetivas do ponto de vista das construções mentais que nós somos capazes de fazer. Kant sustenta que a matemática pura está em um reino de verdades objetivas dependentes da mente e o Intuicionismo matemático de Brouwer é uma outra teoria desse tipo (com a importante diferença de que ela oferece fundamentos para a assertibilidade justificada dos juízos matemáticos em vez de oferecer fundamentos para a verdade deles no sentido clássico). Todas essas posições têm sua natural correspondência na filosofia moral. O intuicionismo do tipo esposado por W. D. Ross talvez seja a posição análoga mais próxima do platonismo matemático e a teoria de Kant é a versão mais familiar da tese de que a moralidade é uma esfera de verdades objetivas que, no entanto, dependem da mente.

Todas essas visões que eu mencionei (com alguma qualificação no caso do convencionalismo) respondem positivamente (isto é, de forma não-cética) à primeira questão filosófica sobre a matemática. Cada uma delas identifica alguns fundamentos objetivos, ou pelo menos intersubjetivos, da verdade para os juízos matemáticos. O ceticismo absoluto e as versões subjetivas das verdades dependentes da mente (análogas ao emotivismo ou ao prescritivismo) são menos atraentes tanto para as filosofias da matemática quanto para as filosofias morais. Isto é, em parte, assim, simplesmente por causa do maior nível de acordo intersubjetivo no juízo matemático. Mas também se deve à diferença das outras questões que as explicações filosóficas dos dois campos precisam responder.

Nem a matemática nem a moralidade podem ser tomadas para descrever o reino de fatos que existem isoladamente do resto da realidade. Supõe-se que elas estão conectadas às outras coisas. Juízos matemáticos dão origem a previsões sobre aqueles reinos para os quais a matemática se aplica. Essa conexão é algo que a explicação filosófica da verdade matemática precisa explicar, mas o fato de que precisamos observar a exatidão de tais previsões e aprender com ela também apoia nossa crença na verdade matemática objetiva. No caso da moralidade, a principal conexão é, ou geralmente supõe que seja, com a vontade. Dado qualquer candidato para o papel sobre o tema da moralidade, nós precisamos explicar por que alguém deveria se preocupar com isso e a necessidade de responder a essa questão de motivação deu forte apoio às visões subjetivistas.

Mas o que uma teoria filosófica adequada sobre moralidade precisa dizer acerca da motivação moral? Ela não precisa, penso eu, demonstrar que a verdade moral dá a alguém que a conheça uma razão para agir que apele aos desejos presentes daquela pessoa ou ao avanço dos interesses dele ou dela. Eu considero totalmente inteligível que a exigência moral possa ser corretamente aplicada a uma pessoa, mesmo se ela não tiver nenhuma razão desse tipo para cumpri-la. Se as exigências morais dão razões para aqueles a quem elas se aplicam em conformidade com algum terceiro elemento, isso é uma questão em disputa que eu vou deixar de lado. Mas o que uma filosofia moral adequada precisa fazer, penso eu, é nos tornar clara a natureza das razões que a moralidade oferece, pelo menos para os que se preocupam com ela. A teoria filosófica da moralidade precisa oferecer uma

explicação sobre aquelas razões que é, de um lado, compatível com sua explicação sobre verdade moral e sobre o raciocínio moral, e é, de outro lado, apoiada por uma análise plausível da experiência moral. Uma filosofia moral satisfatória não se preocupará com a moralidade como uma simples preferência especial, como um fetiche ou um gosto especial, que só a algumas pessoas ocorre ter. Precisa tornar compreensível porque as razões morais são as que as pessoas podem levar a sério e porque atingem aquelas razões que são movidas por elas como razões de um vigor e uma inescapabilidade especiais.

Também há uma outra questão: saber se a susceptibilidade a tais razões é compatível com uma pessoa boa ou se é, como Nietzsche argumentou, um desastre psicológico para a pessoa que a tem. Se alguém deve defender a moralidade, esse alguém precisa demonstrar que ela não é, de nenhuma forma, desastrosa, mas eu não busco aqui essa segunda questão motivacional. Eu a menciono só para distingui-la da primeira questão, que é a minha preocupação atual.

A tarefa de dar uma explicação filosófica do tema da moralidade difere da de analisar o significado dos termos morais e da de encontrar a formulação mais coerente de nossas crenças morais de primeira ordem. Uma ordenação coerente ao máximo de nossas crenças morais de primeira ordem poderia nos fornecer uma espécie valiosa de explicação: ela tornaria claro como vários preceitos, noções e juízos morais, aparentemente disparatados, estão entre si relacionados, indicando, assim, até que ponto os conflitos entre eles são fundamentais e até que ponto, de outro lado, eles podem ser resolvidos ou explicados. Mas a investigação filosófica sobre o tema da moralidade tem uma visão mais externa. Ela procura explicar que tipo de verdade moral as verdades são, descrevendo-as em relação a outras coisas no mundo e em relação às nossas preocupações particulares. Uma explicação de como podemos vir a conhecer a verdade sobre moralidade precisa estar baseada em tal explicação externa do tipo de coisas as verdades morais são, em vez de estar baseada em uma lista de verdades morais particulares, mesmo uma lista coerente ao

máximo. Isso também parece ser verdade em relação às explicações sobre como as crenças morais podem dar a alguém uma razão para agir³.

A coerência entre nossas crenças morais de primeira ordem - o que Rawls chamou de equilíbrio reflexivo estrito⁴ - parece insatisfeita⁵ tanto com a explicação da verdade moral quanto com a explicação da base de justificação na ética, justamente porque, tomada em si mesma, uma explicação coerente ao máximo de nossas crenças morais não precisa nos oferecer aquilo que eu chamei de explicação filosófica sobre o tema da moralidade. Por mais coerentes que possam ser as nossas crenças morais, a dúvida incômoda de que elas não existem, mesmo assim, pode ainda permanecer. Elas podem ser apenas um conjunto de reações socialmente inculcadas, mutuamente consistentes talvez, mas não juízos de um tipo tal a que possa ser dito que são corretos ou incorretos. Uma teoria filosófica sobre a natureza da moralidade pode contribuir para que nós confiemos em nossas crenças morais de primeira ordem, exatamente por dissipar aquelas dúvidas naturais sobre o assunto. Na medida em que ela inclui uma explicação sobre a epistemologia moral, tal teoria pode nos guiar em direção a novas formas de argumento moral, mas ela não precisa fazer isso. Argumento moral do tipo com que nós estamos mais ou menos familiarizados pode permanecer como a única forma de justificação ética. Mas quer isso leve ou não à revisão de nossos modos de justificação, o que uma boa teoria filosófica deveria fazer é nos dar uma compreensão mais clara sobre o que as melhores formas de argumento moral significam e sobre qual tipo de verdade elas são e a que caminhos elas podem levar. (A mesma coisa pode ser dita, eu acredito, sobre a contribuição que a filosofia da matemática faz para a nossa confiança nos juízos matemáticos e nas formas particulares de raciocínio matemático).

³ Embora aqui os laços entre a natureza da moralidade e seu conteúdo sejam mais importantes. Não está claro que uma explicação sobre a natureza da moralidade que deixa seu conteúdo *totalmente* em aberto poderia ser a base para uma explicação plausível sobre a motivação moral.

⁴ Veja Rawls, 1974-1975, p. 08; e Daniels, 1979, p. 257-258. Até que ponto o processo do que eu estou chamando de explicação filosófica coincidirá com a busca pelo “amplo equilíbrio reflexivo”, tal qual é entendido por Rawls e por Daniels é uma outra questão que eu não posso abordar aqui.

⁵ Para expressar essa insatisfação, veja Singer, 1974, e Brandt, 1979, p. 16-21.

Como qualquer tese sobre moralidade, uma explicação filosófica sobre o tema da moralidade precisa ter alguma conexão com o significado dos termos morais: ela precisa ser plausível para afirmar que o assunto descrito é, de fato, ao que esses termos se referem, ao menos, na maioria do uso normal deles. Mas o significado corriqueiro dos termos morais é o produto de muitas crenças morais diferentes defendidas pelos interlocutores passados e presentes da linguagem e esse significado é seguramente compatível com a variedade das visões morais e com a variedade das visões sobre a natureza da moralidade. Afinal, termos morais são usados para expressar muitas visões diferentes desse tipo e as pessoas que expressam essas visões não estão utilizando os termos morais incorretamente, mesmo se o que alguns deles disser deva estar errado. Como um juízo moral de primeira ordem, a caracterização moral do tema da moralidade é uma afirmação substantiva sobre moralidade, porém uma afirmação de um tipo diferente.

Embora a caracterização filosófica da moralidade faça uma espécie de afirmação que difere do juízo moral de primeira ordem, isso não quer dizer que a teoria filosófica da moralidade será neutra em relação às doutrinas normativas que competem entre si. A adoção de uma tese filosófica sobre a natureza da moralidade quase sempre terá algum efeito sobre a plausibilidade das alegações morais em particular, mas as teorias filosóficas sobre moralidade variam amplamente na extensão e na direção das implicações normativas de cada uma delas. Em uma extremidade está o intuicionismo, entendido como tese filosófica segundo a qual a moralidade está preocupada com certas propriedades não-naturais. A retidão, por exemplo, é defendida por Ross⁶ para ser a propriedade que faz o “ajuste” ou “adequação moral”. O intuicionismo defende que nós podemos identificar as ocorrências dessas propriedades e que podemos reconhecê-las como sendo verdades gerais auto-evidentes, mas que elas não podem ser posteriormente analisadas e explicadas em termos de outras noções. Assim compreendido, o intuicionismo é, a princípio, compatível com uma ampla variedade de posições normativas. Poder-se-ia, por exemplo, ser um utilitarista intuicionista ou um crente intuicionista em

⁶ Ross, 1939, p. 52-54, p. 315.

direitos morais, dependendo das verdades gerais sobre a propriedade da retidão moral que se toma para ser auto-evidente.

O outro extremo é representado pelo utilitarismo filosófico. O termo “utilitarismo” é geralmente usado para se referir a uma família de doutrinas normativas específicas - doutrinas que poderiam ser defendidas com base em um conjunto de teses filosóficas diferentes sobre a natureza da moralidade. Nesse sentido do termo, poder-se-ia, por exemplo, ser um utilitarista sobre os fundamentos intuicionista ou contratualista. Mas o que eu vou chamar de “utilitarismo filosófico” é uma tese filosófica particular sobre o tema da moralidade, designadamente a tese de que somente os fatos morais fundamentais são fatos sobre o bem-estar individual⁷. Eu acredito que essa tese tem grande chance de plausibilidade para muitas pessoas, e que, embora algumas pessoas sejam utilitárias por outras razões, ela é o que atrai no utilitarismo filosófico e o que explica a influência generalizada dos princípios utilitaristas.

Parece evidente para as pessoas que há o fato de os indivíduos ficarem em melhor ou pior situação. Tais fatos têm uma força motivacional óbvia; e é bastante compreensível que as pessoas deveriam ser movidas por eles de muitas maneiras, a ponto de supor que elas são movidas pelas considerações morais. Ademais, esses fatos são claramente relevantes para a moralidade como nós agora a compreendemos. Afirmações sobre o bem-estar individual são uma classe de pontos de partida válidos para o argumento moral. Mas a maioria das pessoas acha que é muito mais difícil ver o quanto poderia haver qualquer outro, independente dos pontos de partida. Exigências morais substantivas independentes do bem-estar individual atingem as pessoas como intuicionistas em um sentido censurável. Elas poderiam representar “fatos morais” de um tipo tal que seria difícil explicar. Não há nenhum problema em reconhecer isso como um fato de que um certo ato é, digamos, um caso de mentiras e de quebra de promessa. E um utilitarista pode reconhecer que tais fatos como esses geralmente têm significado moral (derivativo):

⁷ Para os propósitos dessa discussão, eu deixei em aberto questões importantes sobre quais indivíduos devem ser considerados ou como o “bem-estar” deve ser entendido. O utilitarismo filosófico manterá o apelo que me interessa sob muitas respostas diferentes para essas perguntas.

eles são moralmente significativos, por causa de suas consequências para o bem-estar individual. Os problemas, e a acusação de “intuicionismo”, surgem quando se afirma que tais atos estão errados em um sentido que não é redutível ao fato de que eles diminuem o bem-estar individual.

Como essa propriedade independente de equívocos morais poderia ser entendida de uma forma que lhe desse o tipo de importância e força motivacional que as considerações morais foram obrigadas a ter? Se se aceitar a ideia de que não há nenhuma propriedade moral que tenha esse tipo de significado intrínseco, então o utilitarismo filosófico pode parecer que seja a única explicação sustentável sobre moralidade. E uma vez que o utilitarismo filosófico for aceito, algumas formas de utilitarismo normativo parecerão que estão sendo aplicadas a nós como a correta teoria moral de primeira ordem. O utilitarismo então tem, para muitas pessoas, algo como o status que o Formalismo de Hilbert e o Intuicionismo de Brouwer têm para os seus fiéis. É uma visão que parece ser aplicada a nós pela necessidade de dar uma explicação filosoficamente defensável sobre o assunto. Mas ela nos deixa com uma escolha difícil: nós podemos ou abandonar muitas de nossas crenças prévias de primeira ordem ou tentar salvá-las demonstrando que elas podem ser obtidas como verdades derivadas ou explicadas como ficções úteis ou inofensivas.

Aparentemente, o apelo ao utilitarismo filosófico que eu descrevi é falso, uma vez que essa teoria precisa equivaler ou a uma forma de intuicionismo (distinguindo-se de outros somente naquilo que envolve apenas um apelo à intuição) ou a um naturalismo definicional de um outro tipo qualquer refutado por Moore e por outros há muito tempo atrás. Mas eu não penso que a doutrina pode ser descartada assim tão facilmente. O utilitarismo filosófico é uma tese filosófica sobre a natureza da moralidade. Como tal, ele está em pé de igualdade com o intuicionismo ou com a forma de contratualismo que eu depois vou defender neste artigo. Nenhuma dessas teses precisa ser afirmação da verdade como uma questão de definição; se uma delas for verdadeira, não se segue que uma pessoa que a segue estará fazendo mau uso das palavras “certo”, “errado” e “deve”. Nem todas essas teses são formas de intuicionismo, se intuicionismo é entendido como a visão de que os fatos morais se referem especialmente a propriedades não-naturais, que nós podemos apreender pelo discernimento intuitivo, mas que não precisam ou não

admitem qualquer análise posterior. Tanto o contratualismo quanto o utilitarismo filosófico são especialmente incompatíveis com essa afirmação. Como quaisquer outras teses filosóficas sobre a natureza da moralidade (incluindo, eu diria, o próprio intuicionismo), o contratualismo e o utilitarismo filosófico devem ser elogiados com base no sucesso deles em dar uma explicação sobre a crença moral, sobre o argumento moral e sobre a motivação moral que é compatível com nossas crenças gerais sobre o mundo: nossas crenças sobre quais tipos de coisas existem no mundo, sobre quais tipos de observação e de raciocínio de que nós somos capazes e sobre quais tipos de razões nós temos para agir. Um juízo sobre qual explicação da natureza da moralidade (ou da matemática) é mais plausível no sentido geral é só esta: um juízo de plausibilidade geral. Ele não é descrito de forma útil como um discernimento de conceitos ou como um discernimento intuitivo especial de alguma outra espécie.

Se o utilitarismo filosófico for aceito, então alguma forma de utilitarismo aparece para ser aplicada por nós como uma doutrina normativa, mas outro argumento será exigido para determinar qual forma nós deveríamos aceitar. Se tudo que conta moralmente é o bem-estar dos indivíduos, nenhum deles é considerado mais importante do que os outros, e se tudo que importa, no caso de cada indivíduo, é o grau em que o bem-estar dele é afetado, então disso decorreria que a base da avaliação moral tem por o objetivo maximizar a soma⁸ do bem-estar individual. Se esse padrão deve ser aplicado à crítica das ações individuais, ou à seleção de regras ou políticas, ou à inculcação de hábitos e disposições para agir é outra questão, assim como é a questão de como o “bem-estar” em si deve ser compreendido. Portanto, a hipótese de que muito do apelo ao utilitarismo como uma doutrina normativa deriva da atratividade do utilitarismo filosófico explica o quanto as pessoas podem estar convencidas de que alguma forma de utilitarismo deve estar correta, embora já estejam bem incertas quanto à forma pela qual ele se apresenta, se é utilitarismo “direto” ou “na ação” ou alguma forma de utilitarismo indireto “de regra” ou “de motivação”. O que essas visões têm em comum, a despeito de suas

⁸ Atinge-se plausivelmente o “utilitarismo médio” por meio de formas de argumento bem diferentes e algumas delas são mais parecidas com o contratualismo. Eu discuto um desses argumentos na seção IV abaixo.

consequências normativas diferentes, é a identificação da mesma classe de fatos morais fundamentais.

II

Se o que eu disse sobre o apelo ao utilitarismo estiver certo, então o que uma teoria rival precisa fazer é fornecer uma alternativa ao utilitarismo filosófico como concepção do objeto de estudo sobre a moralidade. Isso é o que a teoria que eu vou chamar de contratualismo precisa fazer. Mesmo se isso acontecer e, no entanto, ele for considerado superior ao utilitarismo filosófico como uma explicação sobre a natureza da moralidade, o utilitarismo normativo não terá sido refutado. Restará a possibilidade de que o utilitarismo normativo possa ser estabelecido com outros fundamentos, como, por exemplo, o resultado normativo do próprio contratualismo. Mas um argumento direto e, penso eu, influente para o utilitarismo normativo terá sido deixado de lado.

Dar um exemplo do que eu entendo por contratualismo, uma explicação contratualista sobre a natureza do equívoco moral poderia ser declarada como a seguinte: “Um ato estará errado se a sua execução sob certas circunstâncias não for permitida por qualquer sistema de regras para a regulação geral do comportamento que ninguém poderia racionalmente rejeitar como uma base para o acordo geral informado e voluntário”. Isso é entendido como uma caracterização do tipo de propriedade que o equívoco moral é. Como o utilitarismo filosófico, ele terá consequências normativas, mas não é o meu propósito atual explorá-las detalhadamente. Como uma explicação contratualista de uma noção moral, o que eu apresentei até agora é apenas uma aproximação, que poderia precisar ser consideravelmente modificada. Aqui, eu posso oferecer algumas observações por meio do esclarecimento.

A ideia de “acordo informado” quer dizer excluir o acordo baseado na superstição ou crença falsa sobre as consequências das ações, mesmo se essas crenças fossem tidas por razoáveis para a pessoa em questão. A força que se entende da qualificação “razoável”, por outro lado, é para excluir as rejeições que não seriam razoáveis, dado o objetivo de encontrar princípios que poderiam ser a base do

acordo geral informado e voluntário. Dado esse objetivo, não seria razoável, por exemplo, rejeitar um princípio porque ele impôs um limite a você, quando todo princípio alternativo iria impor limites muito maiores aos outros. Eu terei mais a dizer sobre os fundamentos da rejeição mais à frente no artigo.

A exigência de que o acordo hipotético, que é o objeto do argumento moral, é voluntário significa, não só eliminar a coerção, mas também excluir pessoas forçadas a aceitar um acordo por estarem em uma posição de barganha inferior, por exemplo, porque outras são capazes de aguentar por mais tempo e, por isso, insistir nos melhores termos. O argumento moral é abstraído de tais considerações. A única pressão relevante para o acordo vem do desejo de encontrar e concordar com os princípios de que ninguém que não tivesse esse desejo poderia razoavelmente rejeitar. De acordo com o contratualismo, o argumento moral refere-se à possibilidade de acordo entre pessoas que são todas elas movidas por esse desejo e movidas por ele no mesmo grau. Mas essa suposição contrafactual caracteriza somente o acordo ao qual a moralidade se refere, mas não o mundo ao qual os princípios morais se aplicam. Os que dizem respeito à moralidade, procuram por princípios para a aplicação ao mundo imperfeito deles que eles não poderiam razoavelmente rejeitar e que outros nesse mundo, que agora não são movidos pelo desejo de acordo e que não poderiam razoavelmente rejeitar se viessem, então, a ser movidos⁹.

A explicação contratualista sobre o erro moral se refere a princípios “que ninguém poderia razoavelmente rejeitar” em vez de princípios “que todos poderiam razoavelmente aceitar” pela razão seguinte. Considere um princípio segundo o qual algumas pessoas sofrerão dificuldades severas e suponha que essas dificuldades sejam evitáveis. Isto é, há princípios alternativos segundo os quais ninguém teria de arcar com comparáveis ônus. Poderia acontecer, no entanto, que as pessoas sobre as quais essas dificuldades recaiam estejam em uma determinada situação de autossacrifício e que queiram aceitar esses ônus em prol do que eles veem como o maior bem de todos. Nós diríamos, penso eu, que não seria razoável a eles fazer isso.

⁹ Aqui, eu devo muito a Gilbert Harman pelos comentários que me ajudaram a esclarecer minha declaração sobre o contratualismo.

¹⁰ Um exemplo que devo a Derek Parfit.

Por outro lado, também poderia ser razoável a eles recusar esses ônus e, por isso, poderia não ser razoável para alguém rejeitar um princípio que exige que arque com eles (como não é razoável exigir de alguém que assuma um ônus sem ele querer, da mesma forma, também não é razoável que você exija que alguém deixe um ônus de lado só porque você considera esse ônus irracional). Se essa rejeição fosse razoável, então o princípio que impõe esses ônus é posto em dúvida, apesar do fato de que algumas pessoas em situação particular de autossacrifício pudessem (razoavelmente) aceitá-lo. Portanto, é a razoabilidade de rejeitar um princípio, em vez da razoabilidade de aceitá-lo, que gira em torno do argumento moral.

Parece provável que a maioria dos conjuntos não-equivalentes de princípios passará no teste de não-rejeitabilidade. Isso é sugerido, por exemplo, pelo fato de que há muitos caminhos diferentes para definir deveres importantes, nenhum dos quais é mais ou menos “rejeitáveis” do que os outros. Há, por exemplo, muitos sistemas diferentes de fazer acordo e muitos caminhos diferentes de atribuir responsabilidade para cuidar dos outros. Disso não se segue, no entanto, que qualquer ação permitida por, pelo menos, um desses conjuntos de princípios não possa estar moralmente errado, de acordo com o contratualismo. Se for importante para nós ter algum dever de uma dada espécie (algum dever de fidelidade aos acordos ou algum dever de cuidado mútuo) da qual há muitas formas moralmente aceitáveis, então uma dessas formas precisa ser estabelecida por convenção. Em um conjunto em que uma dessas formas é convencionalmente estabelecidas, ações não permitidas por ela estará errada no sentido da definição dada. Pois, dada a necessidade para tais convenções, uma coisa que não poderia, de forma geral, estar de acordo seria um conjunto de princípios que permitem a alguém desconsiderar definições convencionalmente estabelecidas (e moralmente aceitáveis) de deveres importantes. Essa dependência da convenção introduz um grau de relatividade cultural com a moralidade contratualista. Além disso, o que uma pessoa pode razoavelmente rejeitar dependerá das metas e condições que são importantes para a sua vida e essas também dependerão da sociedade em que ela vive. A definição dada acima permite a variação de ambos os tipos, fazendo com que o erro de uma ação dependa das circunstâncias em que ele é cometido.

A declaração parcial do contratualismo que eu dei tem o caráter abstrato apropriado a uma explicação do objeto da moralidade. Em sua face, ela não envolve nenhuma afirmação específica quanto a quais princípios se poderia estar de acordo ou mesmo se há um único conjunto de princípios que poderia estar na base do acordo. Uma maneira, embora não a única, para o contratualismo chegar às afirmações morais substantivas é dar uma definição técnica da noção relevante do acordo, isto é, especificando as condições segundo as quais o acordo deve ser alcançado, as partes desse acordo e o critério de razoabilidade a ser empregada. Diferentes contratualistas fizeram isso de diferentes formas. O que precisa ser afirmado para tal definição é que (segundo as circunstâncias em que ela é aplicada) o que ela descreve é, de fato, o tipo de acordo voluntário e razoável a que o argumento moral objetiva. Mas o contratualismo também pode ser entendido como uma descrição da matéria objeto da moralidade com base na qual as formas ordinárias do raciocínio moral podem ser entendidas e avaliadas sem proceder por via de uma noção técnica do acordo.

Quem deve ser incluído no acordo geral ao qual o contratualismo se refere? O escopo da moralidade é uma questão difícil de moralidade substantiva, mas uma teoria filosófica da natureza da moralidade deveria fornecer algumas bases para respondê-la. O que uma teoria apropriada deveria fazer é fornecer um quadro dentro do qual o que parece ser argumentos relevantes a favor e contra interpretações particulares do limite moral pode ser realizado. Geralmente, pensa-se que o contratualismo não pode fornecer nenhuma base plausível para uma resposta a essa questão. A crítica acusa ou que o contratualismo não dá nenhuma resposta, porque ele precisa começar com algum conjunto de partes contratantes tomadas como dadas, ou que o contratualismo sugere uma resposta que obviamente é muito restritiva, uma vez que um contrato exige partes que são capazes de fazer e de manter acordos e que são cada uma delas capaz de oferecer aos outros algum benefício em troca de sua cooperação. Nenhuma dessas objeções se aplica à versão do contratualismo que eu defendo. A especificação geral do escopo da moralidade que ele implica, a mim me parece, ser esta: a moralidade se aplica a um ser se a noção da justificação a um ser desse tipo faz sentido. O que se exige para que isso aconteça? Aqui, eu só posso sugerir algumas condições necessárias. A primeira é que

o ser tem um bem, isto é, que haja uma sensação clara na qual se pode dizer que as coisas vão melhor ou pior para esse ser. Isso dá a sensação parcial à ideia do que seria razoável para um gestor aceitar em nome do ser. Seria razoável para um gestor aceitar, pelo menos, essas coisas que são boas, ou não ruins, para o ser em questão. Utilizando essa ideia de gestão, nós podemos estender a noção de aceitação para aplicar aos seres que são incapazes de concordar literalmente com qualquer coisa. Mas essa noção mínima de gestão é muito fraca para dar base à moralidade, de acordo com o contratualismo. A moralidade contratualista conta com noções do que seria razoável aceitar, ou razoável rejeitar, que são essencialmente comparativas. Se não fosse razoável para mim rejeitar um certo princípio, dado o objetivo de encontrar os princípios que ninguém com esse objetivo poderia razoavelmente rejeitar, depende não só de como muitas ações permitidas por esse princípio poderia me prejudicar em termos absolutos, mas também de como essa perda de potencial compara com outras perdas potenciais para outros sob esse princípio e alternativas a ele. Assim, para que um ser mantenha relações morais conosco não é suficiente que tenha um bem, é também necessário que seu bem seja suficientemente similar ao nosso próprio bem para dar base a algum sistema de comparabilidade. Só com base em tal sistema, nós podemos dar um tipo apropriado de sentido à noção do que um gestor poderia razoavelmente rejeitar em nome do ser.

Mas a gama de possíveis gestões é mais ampla do que aquela da moralidade. Alguém poderia agir como um gestor de uma plantação de tomates, de uma floresta ou de uma colônia de formigas e tais entidades não estão incluídas na moralidade. Talvez isso possa ser explicado pelo apelo à exigência de comparabilidade: embora essas entidades tenham um bem, ele não é comparável ao nosso próprio bem de uma forma que dê base a argumentos morais. Ademais, não há, entretanto, nesses casos um ponto de apoio suficiente para a noção da justificação para um ser. Uma outra exigência mínima dessa noção é que o ser constitui um ponto de vista; isto é, que há algo como o que parece ser aquele ser, algo como o que o mundo se parece com ele. Sem isso, nós não permanecemos em uma relação com o ser que faz até a justificação hipotética ser apropriada.

Com base no que eu disse até aqui, o contratualismo pode explicar por que a capacidade de sentir dor deveria se parecer a muitos para contar a favor do status

moral: um ser que tem essa capacidade também parece satisfazer as três condições que eu acabei de mencionar como necessárias à ideia de justificação dele para fazer sentido. Se um ser pode sentir dor, então ele constitui um centro da consciência para a qual a justificação pode ser dirigida. Sentir dor pode ser uma maneira clara de piorar o ser; ter a sua dor aliviada de uma forma que ele possa ser beneficiado; e essas são formas do bem e do mal que parecem diretamente comparáveis com as nossas próprias formas.

Não está claro que as três condições que eu listei como necessárias também são suficientes para a ideia de justificação para um ser fazer sentido. Se elas são ou se elas não são e o que mais pode ser exigido são questões difíceis e controversas. Algo restringiria a esfera moral àqueles cujas justificações poderiam, em princípio, ser comunicadas, ou aos que podem, na realidade, concordarem com algo, ou aos que têm a capacidade de entender o argumento moral. O contratualismo, como eu declarei, não resolve essas questões de uma só vez. Tudo que eu afirmo é que ele dá base ao argumento sobre elas que é, pelo menos, tão plausível quanto aquele oferecido pelas explicações rivais da natureza da moralidade. Essas restrições propostas com o escopo de moralidade são naturalmente entendidas como afirmações discutíveis sobre as condições segundo as quais a noção relevante da justificação faz sentido, e os argumentos comumente oferecidos a favor ou contra elas podem também ser plausivelmente entendidos com esse fundamento.

Algumas outras restrições possíveis com o escopo de moralidade são mais evidentemente rejeitáveis. A moralidade poderia ser restringida para os que têm capacidade de observar suas restrições, ou para os que são capazes de conferir algum benefício futuro recíproco aos outros participantes. Mas é extremamente implausível supor que os seres excluídos por essas exigências estão totalmente fora da proteção da moralidade. O contratualismo como eu o formulei¹¹ pode explicar

¹¹ Sob esse ponto de vista (quando contrastado com alguns outros em que a noção de um contrato é empregado) o que é fundamental para a moralidade é o desejo para acordos razoáveis, não para a busca de vantagens mútuas. Veja a seção V abaixo. Deveria ficar claro que essa versão do contratualismo pode explicar a posição moral de pessoas futuras que será melhor ou pior como resultado do que nós agora estamos fazendo. É menos claro como se poderia lidar com o problema apresentado pelas pessoas futuras que não teriam nascido, mas por nossas ações que também tornaram as condições em que elas vivem piores. Tais pessoas têm razão para rejeitar os princípios que permitem que essas ações sejam

por que isso é assim: só a ausência dessas capacidades não esgota a possibilidade de justificação para um ser. O que ele pode fazer nesses casos, no entanto, é alterar a justificação que são relevantes. Eu sugiro que qualquer que seja a importância que as capacidades para o controle deliberativo e o benefício recíproco podem ter é como fatores que alteram os deveres que os seres têm e os deveres que os outros têm para com eles, não como condições cuja ausência suspende completamente o quadro moral.

III

Eu até agora falei um pouco sobre o conteúdo normativo do contratualismo. Por tudo o que eu disse até aqui, a fórmula utilitária do ato poderia vir a ser um teorema do contratualismo. Não acho que este seja o caso, mas minha tese principal é que qualquer que possa ser as implicações normativas do contratualismo ele ainda tem o conteúdo distinto como uma tese filosófica sobre a natureza da moralidade. Esse conteúdo - a diferença, por exemplo, entre ser um utilitarista porque a fórmula utilitária é a base do acordo geral e ser um utilitarista com outros fundamentos - se mostra mais claramente na resposta que um contratualista dá à primeira pergunta motivacional.

O utilitarismo filosófico é uma visão parcialmente plausível porque os fatos que ele identifica como fundamental para a moralidade - fatos sobre o bem-estar individual - tem uma força motivacional óbvia. Fatos morais podem nos motivar, sob esse ponto de vista, por causa de nossa identificação de simpatia com o bem dos outros. Mas quando nós saímos do utilitarismo filosófico em direção a uma fórmula utilitarista específica como padrão correto para o agir, a forma de motivação a que o utilitarismo apela se torna mais abstrata. Se o utilitarismo clássico for a doutrina normativa correta, então a fonte natural da motivação moral será uma tendência a ir em direção a mudanças no bem-estar agregado, independentemente de como essas mudanças podem ser compostas. Nós precisamos ser motivados da mesma forma por um ganho agregado da mesma magnitude se ele for obtido por aliviar o

realizadas? Esse problema difícil, que eu não posso explorar aqui, foi levantado por Derek Parfit, em *Partif*, 1976.

sofrimento agudo de algumas pessoas ou por trazer pequenos benefícios para um vasto número, talvez às custas de um desconforto moderado para algumas delas. Isso é muito diferente da simpatia do tipo familiar para indivíduos particulares, mas um utilitarista pode argumentar que esse desejo mais abstrato é o que a simpatia natural se torna quando ela é corrigida pela reflexão moral. Esse desejo tem o mesmo conteúdo como simpatia - ele é uma preocupação com o bem dos outros - mas não é parcial ou seletivo em suas escolhas dos objetos.

Deixando de lado a plausibilidade psicológica dessa simpatia imparcial, como poder saber o quanto um candidato é bom para desempenhar o papel de motivação moral? Certamente, a simpatia comum a todos é um dos muitos motivos que, às vezes, podem impelir alguém a fazer a coisa certa. Ela pode ser o motivo determinante, quando, por exemplo, eu corro para ajudar uma criança que está sofrendo. Mas quando eu me vejo convencido pelo artigo de Peter Singer¹² sobre a fome e fico chocado por reconhecer o que parece ser uma exigência moral óbvia, outra coisa entra em ação. Além de pensar sobre o quanto de bem eu poderia fazer às pessoas em locais atingidos pela seca, fico impressionado com o pensamento aparentemente distinto de que seria errado para mim deixar de ajudá-los, mesmo quando eu poderia ajudar, a um custo muito baixo para a minha realidade. Um utilitarista pode responder que sua explicação de motivação moral não pode falhar por não captar esse aspecto da experiência moral, uma vez que ela é apenas uma reflexão de nossa educação moral não-utilitária. Além disso, ela não deve ter fundamento. Para qual tipo de fato esse outro suposto fato de erro moral poderia acontecer, e como ele poderia nos dar uma outra razão especial para agir? A pergunta em relação ao contratualismo, então, é se ele pode dar uma resposta satisfatória para esse desafio.

De acordo com o contratualismo, a fonte de motivação que é diretamente provocada pela crença de que uma ação está errada é o desejo de ser capaz de justificar as ações de alguém para os outros sob o fundamento de que eles não

¹² Singer, 1972.

podem rejeitar razoavelmente¹³. Eu acho isso uma explicação extremamente plausível de motivação moral - pelo menos, uma melhor explicação de minha experiência moral do que a alternativa utilitária natural - e me parece constituir um ponto forte para a visão contratualista. Todos nós poderíamos gostar de, realmente, estar de acordo com as pessoas ao nosso redor, mas o desejo que o contratualismo identifica como básico para a moralidade não nos leva simplesmente a conformar com os padrões aceitos pelos outros, independentemente do que esses padrões possam ser. O desejo de ser capaz de justificar as ações de alguém para os outros sob o fundamento de que eles não podem rejeitar com razoabilidade será satisfeito quando soubermos que há uma justificação adequada para nossa ação, ainda que os outros, de fato, recusem aceitá-la (talvez porque eles não tenham nenhum interesse em achar princípios que nós e os outros não pudessem razoavelmente rejeitar). Do mesmo modo, uma pessoa movida por esse desejo não se satisfará pelo fato de que os outros aceitem uma justificação para a ação dela, se ela considera essa justificação como espúria.

Um bom teste para se saber se você considera uma justificação suficiente é se você aceitaria essa justificação se você estivesse em uma posição de outra pessoa. Essa conexão entre a ideia de “mudança de lugar” e a motivação subjacente à moralidade explica a ocorrência frequente do argumento da “Regra de Ouro” dentro de diferentes sistemas de moralidade e no ensino de várias religiões. Mas o experimento de pensamento da mudança de lugar é apenas um bom guia; a pergunta fundamental é o que seria irrazoável para rejeitar como base para acordos informados, não forçados e gerais. Conforme a observação feita por Kant¹⁴, nossos diferentes pontos de vista individuais, tomados como eles são, podem, de um modo geral, ser simplesmente irreconciliáveis. “Harmonia de juízos” exige a construção de uma forma genuinamente interpessoal de justificação que, não obstante, seja algo com que cada indivíduo pudesse concordar. Desse ponto de vista interpessoal, uma certa quantidade de como as coisas se parecem do ponto de vista de outra pessoa,

¹³ Razoavelmente, isto é, dado o desejo de achar os princípios que os outros, da mesma forma motivados, não poderiam rejeitar razoavelmente.

¹⁴ Kant, 1785, seção 02, nota de rodapé 14.

como uma certa quantidade de como elas se parecem do meu próprio ponto de vista, será contada como um viés.

Eu não estou afirmando que o desejo de ser capaz de justificar as ações de alguém para com outros sob o fundamento de que elas não poderiam razoavelmente rejeitar é universal ou “natural”. A “educação moral” me parece plausivelmente entendida como um processo de cultivar esse desejo e de moldá-lo, em grande parte aprendendo o que as outras justificações estão, de fato, dispostos a aceitar, descobrindo quais você mesmo considera aceitáveis quando confrontá-los com uma variedade de perspectivas e avaliando se você mesmo e os outros aceitam ou rejeitam essas justificações à luz de uma experiência maior.

Na verdade, parece-me que o desejo de ser capaz de justificar as próprias ações (e instituições) em bases que consideramos aceitáveis é bastante forte na maioria das pessoas. As pessoas estão dispostas a percorrer distâncias consideráveis, que envolvem sacrifícios muito pesados, a fim de evitar admitir a injustificabilidade de suas ações e instituições. A notória insuficiência da motivação moral como meio de convencer as pessoas a fazer a coisa certa não se deve simplesmente à fraqueza do motivo subjacente, mas também ao fato de que é facilmente desviado pelo auto-interesse e pela autodecepção.

Poderia ser razoável objetar aqui que a fonte da motivação que eu descrevi não está exclusivamente atada à noção contratualista de verdade moral. A explicação da motivação moral que eu ofereci se refere à ideia de uma justificação que seria irrazoável rejeitar e essa ideia é potencialmente mais ampla do que a noção contratualista de acordo. Pois permite que M seja alguma explicação não-contratualista de verdade moral. De acordo com M, podemos supor que o equívoco da uma ação é simplesmente uma característica daquela ação em virtude da qual ela não deve ser feita. Uma ação que tem essa característica, de acordo com M, a tem muito, independentemente de qualquer tendência de pessoas informadas que venham a concordar com ela. No entanto, uma vez que as pessoas informadas estejam presumivelmente em uma posição para reconhecer o equívoco de um tipo na ação, seguir-se-ia daí que se uma ação estivesse errada, então tais pessoas concordariam que ela não seria realizada. Do mesmo modo, se um ato não é moralmente errado e há uma justificação moral adequada para realizá-lo, então

presumivelmente haverá uma justificação moral para ele que a uma pessoa informada seria irrazoável rejeitar. Portanto, mesmo se M, e não o contratualismo, for a explicação correta da verdade moral, o desejo de ser capaz de justificar minhas ações aos outros, com o fundamento de que eles não poderiam razoavelmente rejeitar, poderia ainda servir como base para a motivação moral.

O que isso mostra é que o apelo tanto ao contratualismo quanto ao utilitarismo repousa, em parte, em um ceticismo qualificado. Uma teoria não-contratualista da moralidade pode fazer uso da fonte de motivação à qual o contratualismo apela. Mas um argumento moral recorrerá a essa fonte de motivação justamente porque ela é uma boa justificação para agir de um certo modo, uma justificação que a outros não seria irrazoável aceitar. Portanto, uma teoria não-contratualista precisa dizer que há propriedades morais que têm força justificatória bastante independente de seu reconhecimento de qualquer acordo ideal. Essas representariam o que John Mackie chamou de instâncias do “a-ser-feito” e do “não-a-ser-feito” intrínsecos¹⁵. Parte do apelo contratualista repousa na visão de que, segundo o ensinamento de Mackie, é intrigante como poderia haver tais propriedades “no mundo”. Em contrapartida, o contratualismo procura explicar o status justificatório das propriedades morais, bem como sua força motivacional, em termos da noção de acordo razoável. Em alguns casos, as propriedades morais são em si para ser entendidas em termos dessa noção. Isso é assim, por exemplo, no caso da propriedade do erro moral, acima considerado. Mas também há propriedades de fazer o bem e o mal que são em si independentes da noção contratualista do acordo. Eu considero a propriedade de ser um ato de matar pelo prazer de fazer assim para ser uma propriedade de fazer o mal desse tipo. Tais propriedades são fazer o mal, porque seria razoável rejeitar qualquer conjunto de princípios que permitiram os atos que elas caracterizam. Assim, embora haja propriedades moralmente relevantes “no mundo” que são independentes da noção contratualista do acordo, essas não constituem instâncias do “a-ser-feito”: a sua relevância moral - sua força nas justificações bem como sua ligação com a motivação - deve ser explicada em fundamentos contratualistas.

¹⁵ Mackie 1977, p. 42.

Em particular, o contratualismo pode explicar o aparente significado moral dos fatos sobre o bem-estar individual, que o utilitarismo considera ser fundamental. O bem-estar individual será moralmente significativo, de acordo com o contratualismo, não porque é intrinsecamente valoroso ou porque promovê-lo é auto-evidentemente uma característica de fazer o bem, mas simplesmente porque um indivíduo poderia razoavelmente rejeitar uma forma de argumento que não deu nenhum peso ao seu bem-estar. Essa afirmação de significado moral é, no entanto, apenas aproximada, uma vez que é exatamente uma questão difícil como “bem-estar” ser entendida e de que maneira somos obrigados a levar em conta o bem-estar dos outros ao decidir o que fazer. Não se segue dessa afirmação, por exemplo, que um certo desejo terá sempre e em todo lugar o mesmo peso ao determinar a correção de uma ação que promoveria sua satisfação, um peso proporcional à sua força e “intensidade”. A força para fazer o bem dos desejos de uma pessoa é especificada pelo que poderia ser chamada de concepção de interesses moralmente legitimados. Tal concepção é um produto do argumento moral; isso não está dado, como a noção de bem-estar individual pode estar, simplesmente pela ideia do que é racional para um indivíduo desejar. Nem tudo para o qual eu tenho um desejo racional será algo em que outros precisam me conceder para ter um interesse legítimo que eles se comprometeram a ponderar para decidir o que fazer. Uma variedade de coisas que podem ser objetos dos meus desejos racionais é muito amplo de fato, e a variedade de afirmações que outros não poderiam razoavelmente recusar a reconhecer quase certamente será mais estrita do que isso. Haverá uma tendência pelos interesses para se conformar ao desejo racional - àquelas condições que o tornam racional para desejar algo também para estabelecer um interesse legítimo nele - mas os dois nem sempre coincidirão.

Um efeito do contratualismo, portanto, é quebrar a distinção rígida, que argumenta a favor do apelo utilitarista, entre o status de bem-estar individual e o das outras noções morais. Exige-se um quadro do argumento moral para definir nossos interesses legítimos e para explicar a sua força moral. Esse mesmo quadro contratualista também pode explicar a força das outras noções morais como, por exemplo, direitos, responsabilidade individual e equidade procedimental.

IV

Parece improvável que o utilitarismo na ação seja um teorema da versão do contratualismo que eu descrevi. O significado moral positivo dos interesses individuais é uma reflexão direta sobre a exigência do contratualismo de que as ações sejam defensáveis a cada pessoa com os fundamentos de que ela não poderia razoavelmente rejeitar. Mas é um longo passo daqui até a conclusão com que cada indivíduo deve concordar para sempre deliberar do ponto de vista do benefício agregado máximo e para aceitar justificações, só apelando para essa consideração. É bem possível que, de acordo com o contratualismo, algumas questões morais podem ser adequadamente resolvidas apelando ao bem-estar agregado máximo, mesmo se isso não for o único ou último padrão de justificação.

O que parece mesmo improvável é que o contratualismo deveria acabar coincidindo com alguma forma de utilitarismo de “dois níveis”. Eu não posso avaliar toda essa possibilidade aqui. O contratualismo não compartilha com essas teorias as características importantes que a defesa das ações individuais deve se proceder por meio de uma defesa de princípios que permitiriam essas ações. Mas o contratualismo difere de algumas formas do utilitarismo de dois níveis de um jeito importante. O papel dos princípios no contratualismo é fundamental; eles não entram simplesmente como mecanismos para a promoção do agir que são corretos de acordo com algum outro padrão. Uma vez que ele não estabeleça duas formas potencialmente conflituosas de raciocínio moral, o contratualismo evita a instabilidade que geralmente assola o utilitarismo de regras.

A pergunta fundamental aqui, no entanto, é se os princípios para os quais o contratualismo leva devem ser aqueles cuja adoção geral (idealmente ou segundo algumas condições mais realistas) promoveria o bem-estar agregado máximo. Pareceu a muitos que esse deve ser o caso. Para sinalizar que eu não concordo, eu vou considerar um dos melhores e mais conhecidos argumentos para essa conclusão e explicar por que eu não penso que ele seja exitoso. Isso também ensejará uma oportunidade para examinar a relação entre a versão do contratualismo que eu defendi aqui e a versão apresentada por Rawls.

O argumento que eu vou considerar, que é familiar aos textos de Harsanyi¹⁶ e outros, procede por meio de uma interpretação da noção contratualista da aceitação e leva ao princípio da utilidade média máxima. Pensar sobre um princípio como candidato a acordo unânime, no qual eu devo pensar sobre ele, não simplesmente como algo aceitável para mim (talvez em virtude de minha posição particular, meus gostos etc.), mas como aceitável¹⁷ para os outros também. Para ser relevante, o meu juízo segundo o qual o princípio é aceitável deve ser imparcial. Mas o que isso significa? Julgar imparcialmente que um princípio é aceitável é, poder-se-ia dizer, julgar que é um princípio que você teria motivos para aceitar não importando quem você fosse. Isto é, e aqui está a interpretação, julgar que é um princípio que seria racional aceitar se você não soubesse qual posição pessoal você ocuparia e acreditasse que você tivesse igual chance de estar em qualquer dessas posições. (“Estar na posição de uma pessoa aqui é entendido e significa estar nas circunstâncias objetivas dela e avaliá-las a partir da perspectiva dos gostos e das preferências dela). Mas, reitere-se, o princípio que seria racional preferir segundo essas circunstâncias - o princípio que ofereceria utilidade de maior expectativa para quem escolhe - seria aquele princípio segundo o qual a utilidade média das partes afetadas fosse a mais alta.

Esse argumento poderia ser questionado em inúmeros pontos, mas o que me preocupa no presente é a interpretação de imparcialidade. O argumento pode ser derrubado em três estágios. O primeiro deles é a ideia de que os princípios morais devem ser aceitáveis de forma imparcial. O segundo é a ideia de escolher os princípios ignorando a posição de alguém (incluindo os seus gostos, preferências etc.). O terceiro é a ideia de escolha racional, partindo da suposição de que alguém tem a mesma chance de ocupar a posição de outra pessoa. Vou deixar de lado por um momento a passagem do estágio-dois para o estágio-três, e me concentrar no

¹⁶ Veja Harsanyi, 1955, sec. IV. Aqui ele está discutindo um argumento que ele apresentou antes em Harsanyi 1953.

¹⁷ Quando eu discuto Harsanyi e Rawls, geralmente os seguirei ao falar da aceitabilidade dos princípios em vez da não-rejeitabilidade deles. A diferença entre elas, acima apontada, é importante somente dentro da versão do contratualismo que eu estou apresentando; assim sendo, eu vou falar da rejeitabilidade somente quando eu estiver contrastando minha própria versão com as deles.

primeiro passo, do estágio-um para o estágio-dois. Há um meio de fazer as coisas parecido com esse passo que é, penso eu, bastante válido, mas não se chega à conclusão necessária ao argumento. Se eu acreditar que um certo princípio P não poderia ser razoavelmente rejeitado como uma base para um acordo geral informado e não forçado, então eu terei de acreditar não só que é algo que seria razoável para mim aceitar, mas algo que seria razoável para os outros também aceitarem, na medida em que todos nós estamos procurando um fundamento para o acordo geral. Consequentemente, eu devo acreditar que eu teria motivos para aceitar P não importando qual posição social eu estivesse ocupando (no entanto, pelos motivos acima mencionados, eu não posso acreditar que eu concordaria com P se eu estivesse em alguma dessas posições). Agora, seria possível pensar que nenhum sentido pode ser anexado à noção de escolher ou concordar com um princípio, ignorando a posição social de alguém, especialmente quando isso inclui ignorar os gostos, as preferências, etc., de alguém. Mas há ao menos um sentido mínimo que poderia ser anexado a essa noção. Se fosse razoável para todos escolher ou concordar com P, então meu conhecimento de que eu tenho motivos para fazer assim não precisa depender do meu conhecimento sobre a minha posição particular, dos meus gostos, das minhas preferências etc. Portanto, na medida em que faz todo e qualquer sentido falar da escolha de e do acordo com algo na ausência desse conhecimento, poder-se-ia dizer que eu tenho motivos para escolher ou concordar com aquelas coisas com que todos têm motivos para escolher ou concordar (aceitando, de novo, o objetivo de encontrar princípios com os quais todos poderiam concordar). E de fato, esse mesmo raciocínio pode nos levar a uma versão do estágio-três. Pois se eu julgo que P é um princípio que todos tem motivos para concordar, então se poderia dizer que eu teria motivos para concordar com ele, se eu pensasse que eu tivesse igual chance de ser alguém ou, de fato, se eu atribuísse a qualquer outra pessoa um conjunto de probabilidades para ser uma ou outra das pessoas em questão.

Mas fica claro que isso não é a conclusão a que o argumento original visou. Essa conclusão se preocupou com o que seria racional para uma pessoa auto-interessada em escolher ou concordar, partindo-se da suposição da ignorância ou da igual probabilidade de ser alguém. A conclusão a que nós chegamos apela para uma

noção diferente: a ideia do que seria irrazoável às pessoas rejeitarem, dado que elas estão procurando uma base para o acordo geral. A direção da explicação nos dois argumentos é bastante diferente. O argumento original costumava explicar a noção de aceitabilidade imparcial de um princípio ético apelando à noção de escolha racional auto-interessada segundo condições especiais, uma noção que parece ser uma noção mais clara. Meu argumento revisado explica como um sentido poderia ser anexado à ideia de escolha ou acordo ignorando a posição de alguém, dada alguma ideia do que seria irrazoável para alguém rejeitar como base para o acordo geral. Isso indica um problema para minha versão do contratualismo: ela poderia ser acusada de falhar na explicação da noção central de que ela é dependente. Aqui, eu responderia que minha versão do contratualismo não procura explicar essa noção. Ela apenas tenta descrevê-la claramente e mostrar como outras características da moralidade podem ser entendidas nos termos dela. Em particular, ela não tenta explicar essa noção reduzindo-a à ideia do que maximizaria as expectativas auto-interessadas de uma pessoa, se ela estivesse escolhendo a partir de uma posição de ignorância ou segundo a suposição da igual probabilidade de ser qualquer pessoa.

386

A plausibilidade inicial da mudança do estágio-um para o estágio-dois do argumento original repousa sobre uma sutil transição de uma dessas noções para a outra. Quem acredita que um princípio é um princípio moralmente correto deve acreditar que ele é um princípio com que todos poderiam razoavelmente concordar e ninguém poderia razoavelmente rejeitar. Mas a minha crença de que isso seja o caso pode geralmente ser distorcida por uma tendência de tomar para mim a sua vantagem de forma mais grave do que seus possíveis custos aos outros. Por essa razão, a ideia de “me pôr no lugar de uma outra pessoa” é um mecanismo de correção útil. O mesmo pode ser dito para o experimento pensado para responder com o que eu poderia concordar ignorando a minha posição. Mas dois desses experimentos pensados são mecanismos para considerar mais agudamente a questão com a qual todos poderiam razoavelmente concordar ou o que ninguém poderia razoavelmente rejeitar. Isto é, eles envolvem o padrão de raciocínio exibido em minha forma revisada do argumento em três estágios, não aquela do argumento como originalmente dado. A pergunta, o que maximizaria as expectativas de uma pessoa auto-interessada que escolhe ignorando sua verdadeira posição, é uma

pergunta bastante diferente. Isso pode ser visto considerando a possibilidade de que a distribuição com a utilidade média mais alta, chamem-na de A, poderia envolver extremamente níveis de utilidade baixas para algumas pessoas, níveis muito mais baixos do que qualquer outro nível mínimo experimental, segundo uma distribuição mais igual.

Vamos supor que A fosse um princípio que seria racional para o auto-interesse de quem escolhe com igual chance de estar na posição de qualquer pessoa para fazer a seleção das coisas. Disso se segue que ninguém poderia razoavelmente rejeitar A? Parece evidente que isso não se segue¹⁸. Vamos supor que a situação dos que se saem pior sob A, chamemo-los de Perdedores, é extremamente ruim e que há uma alternativa para A, chamemo-la de E, sob a qual a situação de ninguém sequer seria tão ruim quanto essa. Prima facie, os perdedores pareceriam ter um fundamento razoável para reclamar contra A. A objeção deles pode ser refutada, apelando aos sacrifícios que seriam impostos a algum outro indivíduo pela seleção de E em vez de A. Mas o mero fato de que A produz utilidade média mais elevada, que poderia ser atribuída ao fato de que muitas pessoas se saem ligeiramente melhor sob A do que sob E, embora algumas delas se saiam pior, não resolve a questão.

Segundo o contratualismo, quando consideramos um princípio, nossa atenção naturalmente se dirige primeiro para os que se saem pior sob ele. Isso ocorre porque, se há pessoas que têm fundamentos razoáveis para se opor ao princípio, é provável que sejam eles. Não se segue daí, no entanto, que o contratualismo sempre exija que nós selecionemos o princípio segundo o qual as expectativas dos que estão em pior situação sejam as mais elevadas. A razoabilidade da objeção dos Perdedores a A não é estabelecida simplesmente pelo fato de que eles estejam em pior condição sob A e de que ninguém estaria nessa condição ruim sob E. Essa reclamação deve ser ponderada com aquelas dos indivíduos que se sairiam pior sob E. A pergunta a ser respondida é: é irracional para alguém recusar a tolerar a situação dos Perdedores sob A, a fim de que alguma outra pessoa devesse ser capaz de desfrutar dos benefícios de que ela teria de renunciar sob E? À medida que a suposta situação do

¹⁸ A discussão que se segue tem muito em comum com o contraste entre princípios majoritários e princípios unânimes esboçados por Thomas Nagel em “Igualdade”, capítulo 08, de Nagel, 1979. Sou grato à discussão de Nagel a respeito dessa ideia.

Perdedor sob A torna-se melhor, ou seu ganho sob E diminui em relação aos sacrifícios exigidos para produzi-lo, sua alternativa se enfraquece.

Uma característica digna de nota do argumento contratualista como eu a apresentei até agora é que ela é não-agregadora: o que se compara são as perdas e os ganhos individuais, bem como os níveis de bem-estar. O quanto as considerações agregadoras podem fazer parte do argumento contratualista é uma outra questão muito mais ampla para ser discutida aqui.

Eu tenho criticado um argumento a favor do Utilitarismo Médio que geralmente está associado a Harsanyi, sendo que minhas objeções a esse argumento (deixando de lado as últimas observações sobre o maximin) têm uma óbvia similaridade com as objeções levantadas por Rawls¹⁹. Mas as objeções que eu levantei também se aplicam contra algumas características do próprio argumento de Rawls. Rawls aceita o primeiro passo do argumento que eu descrevi. Isto é, ele acredita que os princípios corretos de justiça são os que “as pessoas racionais preocupadas em promover seus interesses” aceitariam segundo as condições definidas por sua Posição Original, onde elas seriam ignorantes de seus próprios talentos particulares, sua concepção do bem e sua posição social (ou geração) na qual elas nasceram. É o segundo passo do argumento que Rawls rejeita, isto é, a afirmação de que seria racional às pessoas então situadas para escolher aqueles princípios que lhes ofereceriam a maior utilidade esperada segundo a suposição de que elas têm igual chance de ser alguém na sociedade em questão. Eu acredito, no entanto, que um erro já foi cometido, quando o primeiro passo foi dado.

Isso pode ser evidenciado considerando a ambiguidade na ideia de aceitação pelas pessoas “preocupadas em promover seus interesses”. Em uma leitura, esse é um ingrediente essencial no argumento contratual; em outra, ele é evitável e, penso eu, errado. Na primeira leitura, os interesses em questão são simplesmente aqueles dos membros da sociedade para quem os princípios de justiça devem ser aplicados (e por quem aqueles princípios devem, em última análise, ser aceitos). O fato de que eles têm interesses que podem conflitar, e que eles também estão preocupados em

¹⁹ Por exemplo, o argumento intuitivo contra o utilitarismo na página 14 de Rawls, 1971, e sua observação repetida de que nós não podemos esperar que algumas pessoas aceitem padrões inferiores de vida em prol de expectativas mais elevadas de outras.

promover, é o que dá substância às questões de justiça. Na segunda leitura, a preocupação “em promover seus interesses” que está em questão é uma preocupação das partes com a Posição Original de Rawls, e é essa preocupação que determina, em primeira instância²⁰, quais princípios de justiça que elas adotarão. O acordo unânime entre essas partes, cada qual motivada a fazer para si o máximo de bem que puder, deve ser obtido privando-as de qualquer informação que pudessem dar a elas razões para escolher de forma diferente uma da outra. Por trás do véu da ignorância, o que oferece as melhores perspectivas para alguém, também oferecerá a melhor perspectiva para todos, uma vez que ninguém pode dizer o que o beneficiaria em particular. Então, a escolha de princípios pode ser feita, segundo Rawls, a partir do ponto de vista de uma simples indivíduo racional atrás do véu da ignorância. Quaisquer que sejam as regras de escolha racional que é dito para esse simples indivíduo, preocupado em promover seus próprios interesses da melhor forma que ele pode promover, empregar, essa redução do problema ao caso de uma escolha auto-interessada de uma simples pessoa deveria levantar nossas suspeitas. Conforme eu já havia demonstrado na crítica a Harsanyi, é importante perguntar se a esse simples indivíduo é assegurado aceitar um princípio porque ele julga que é o único princípio que o indivíduo não poderia razoavelmente rejeitar pouco importando a posição que ele viesse a ocupar, ou se, ao contrário, se supõe que o princípio pode ser aceito por uma pessoa em qualquer posição social porque ele seria a escolha racional para uma simples pessoa auto-interessada atrás do véu da ignorância. Eu sustentei acima que o argumento em favor do utilitarismo médio envolve uma transição encoberta do primeiro padrão de raciocínio para o segundo. O argumento de Rawls também parece ser dessa segunda forma; a defesa dele de seus dois princípios de justiça conta, pelo menos inicialmente, com afirmações sobre o que seria racional para uma pessoa, preocupada em promover seus próprios interesses, escolher atrás do véu da ignorância. Eu afirmaria, no entanto, que a plausibilidade dos argumentos de Rawls que favorecem seus dois princípios em

²⁰ Embora elas possam, então, checar para ver que os princípios que elas escolheram serão estáveis, não produz tensões intoleráveis no compromisso e assim por diante. Como eu argumentei acima, essas outras considerações podem ser interpretadas de um jeito que traga a teoria de Rawls para mais perto da versão do contratualismo aqui apresentada.

relação ao princípio da utilidade média está preservada e, em alguns casos aprimorada, quando eles são interpretados como instâncias da primeira forma de argumento contratualista.

Alguns desses argumentos têm um caráter moral informal. Eu já mencionei a observação dele sobre a inaceitabilidade de impor expectativas menores de alguns em prol das expectativas maiores de outros. Mais especificamente, ele diz para as partes na Posição Original que estão preocupadas “em escolher princípios cujas consequências estão preparadas para viver com qualquer geração a que elas vierem a pertencer”²¹ ou, presumidamente, com qualquer posição social com que elas vierem a estar. Essa é uma declaração clara da primeira forma do argumento contratualista. Um pouco mais à frente, ele observa, em favor dos dois princípios, que ambos “são aqueles que uma pessoa escolheria para o projeto de uma sociedade em que se deve atribuir um lugar para o inimigo dela”²². Rawls continua rejeitando essa observação, dizendo que as partes “não deveriam raciocinar a partir de premissas falsas”²³, mas vale a pena perguntar por que pareceu uma coisa plausível dizer em primeiro lugar. A razão, presumo, é essa. Em um argumento contratualista da primeira forma, cujo objeto é encontrar princípios aceitáveis para cada pessoa, atribuição de um oponente malévolo é um experimento pensado que tem um papel heurístico igual àquele do véu da ignorância: é um meio de testar se alguém realmente julga um princípio para ser aceitável a partir de todos os pontos de vista ou se, ao contrário, alguém está falhando em levar a sério o efeito dele às pessoas em posições sociais outras do que as de alguém especificamente.

Mas essas são, todas elas, observações informais e é correto supor que o argumento de Rawls, como o argumento em favor da utilidade média, tem a intenção de se sair da ideia contratualista informal de princípios “aceitáveis a todos” para a ideia de escolha racional por trás do véu da ignorância, uma ideia que é, ele espera, mais precisa e mais capaz de produzir resultados definitivos. Deixe-me voltar então aos seus argumentos mais formais em favor da escolha do Princípio da

²¹ Rawls, 1971, p. 137.

²² Rawls, 1971, p. 152.

²³ Rawls, 1971, p. 137.

Diferença pelas partes na Posição Original. Rawls cita três características da decisão enfrentada pelas partes na Posição Original que, afirma ele, a torna racional para que elas usem a regra maximin e, assim, escolham seu Princípio da Diferença como um princípio de justiça. Essas três características são: (1) a ausência de qualquer base objetiva para estimar probabilidades, (2) o fato de que alguns princípios poderiam ter consequências para os que “eles dificilmente poderiam aceitar”, embora (3) seja possível para eles (seguindo a maximin) assegurar-se de uma perspectiva mínima, avanços acima dos quais, comparativamente, importam muito pouco²⁴. A primeira dessas características é ligeiramente complicada e eu a deixo de lado. Parece claro, no entanto, que as outras considerações mencionadas têm, pelo menos, força em um argumento contratualista informal sobre o qual todos poderiam razoavelmente concordar, tanto quanto elas agem para determinar a escolha racional de uma pessoa simples preocupada em promover seus interesses. Elas expressam a força da objeção que os “perdedores” poderiam ter para um esquema que maximizasse a utilidade média às suas custas, quando comparada com as contra-objeções de que os outros poderiam ter para um arranjo mais igualitário.

391

Em acréscimo a esse argumento sobre a escolha racional, Rawls invoca, entre “os principais fundamentos para os dois princípios”, outras considerações que, como eles dizem, usam o conceito de contrato em maior medida²⁵. As partes na Posição Original, diz Rawls, podem concordar com os princípios de justiça somente se eles pensam que esse acordo é o acordo que eles, na verdade, serão capazes de cumprir. É, ele afirma, mais plausível acreditar nesses dois princípios do que no princípio da utilidade média, segundo o qual os sacrifícios demandados (“as tensões do compromisso”) poderão ser muito mais elevados. Um segundo, a afirmação relatada, é que os dois princípios de justiça têm maior estabilidade psicológica do que o princípio da utilidade média. É mais plausível acreditar, afirma Rawls, que, em uma sociedade em que eles foram realizados, as pessoas continuariam a aceitá-los e a ser motivadas a agir de acordo com eles. A aceitação contínua do princípio da utilidade

²⁴ Rawls 1971, p. 154.

²⁵ Rawls 1971, seq. 29, p. 175ff.

média, por outro lado, exigiria um grau excepcional de identificação com o bem por parte daqueles de quem os sacrifícios foram demandados.

Essas observações podem ser entendidas como afirmações sobre a “estabilidade” (em um sentido bastante prático) de uma sociedade fundada nos dois princípios de Rawls sobre justiça. Mas eles também podem ser vistos como uma tentativa de mostrar que um princípio vindo por meio da segunda forma de raciocínio contratualista também satisfará as exigências da primeira forma, isto é, aquilo que é algo que ninguém poderia razoavelmente rejeitar. A questão “A aceitação desse princípio é um acordo que você, na verdade, poderia realizar?” é, como a ideia da atribuição do pior inimigo de alguém, um experimento mental pelo qual podemos usar nossas próprias reações para testar nosso juízo de que certos princípios são princípios que alguém poderia razoavelmente rejeitar. Princípios gerais da psicologia humana também podem ser invocados para este mesmo fim.

O argumento final de Rawls é que a adoção de seus dois princípios dá apoio público ao autorrespeito dos membros individuais da sociedade e “dá uma interpretação mais forte e mais característica da ideia de Kant²⁶” de que as pessoas devem ser tratadas como fins, não simplesmente como meios para o maior bem coletivo. Mas, qualquer diferença que possa existir aqui entre os dois princípios de justiça em Rawls e o princípio de utilidade média, há, pelo menos nitidamente, um contraste entre os dois padrões de raciocínio contratualista acima distinguidos. A conexão com o autorrespeito e com a fórmula kantiana é preservada pela exigência de que os princípios da justiça sejam aqueles que nenhum membro da sociedade poderia razoavelmente rejeitar. Essa conexão se enfraquece quando alteramos a ideia de uma escolha que promove os interesses de um simples indivíduo racional para quem as várias vidas individuais em uma sociedade são, justamente, muitas possibilidades diferentes. Isso é assim qualquer que seja a regra de decisão que se diz para empregar àquele que vai racionalmente escolher. O argumento do maximin parece preservar essa ligação porque reproduz, como uma afirmação sobre a escolha racional, o que é, em termos ligeiramente diferentes, um argumento moral atraente.

²⁶ Rawls, 1971, p. 183.

A “situação de escolha” que é fundamental para o contratualismo conforme eu o descrevi é obtida começando com os indivíduos “mutuamente desinteressados” com o pleno conhecimento de suas situações e, acrescentado a isso (não, como às vezes se sugere, benevolência, mas), um desejo a cada uma de suas partes para encontrar princípios que ninguém poderia razoavelmente rejeitar, na medida em que eles também têm esse desejo. Rawls inúmeras vezes considera, de passagem, tal ideia²⁷. Ele a rejeita em favor de sua própria ideia de escolha mutuamente desinteressada por trás do véu da ignorância com o fundamento de que somente esta última nos capacita a chegar aos resultados definitivos: “Se ao escolher princípios de que nós exigimos unanimidade mesmo onde há completa informação, apenas alguns casos bastante óbvios poderiam ser decididos”²⁸. Eu acredito que essa suposta vantagem é questionável. Talvez porque minhas expectativas para o argumento moral sejam interpretadas como argumentos dentro da forma do contratualismo que eu estava propondo. Uma possível exceção é o argumento a partir do maximin. Se o Princípio da Diferença fosse tomado para ser, de um modo geral, aplicável às decisões de política pública, então a segunda forma de raciocínio contratualista, por meio da qual ela deriva, teria implicações de maior alcance do que a forma do argumento do perdedor, em comparação aos perdedores, que eu empreguei. Mas essas implicações mais amplas do princípio não são sempre plausíveis e eu não acho que Rawls o compreenda para ser aplicado tão amplamente. Sua intenção é que o Princípio da Diferença deve ser aplicado só para as iniquidades maiores geradas pelas instituições básicas de uma sociedade, e essa limitação é uma reflexão sobre as condições especiais segundo as quais ele defende o maximin para ser a base apropriada para a escolha racional: algumas escolhas produzem resultados que alguém dificilmente poderia aceitar, enquanto ganhos acima do mínimo que se pode assegurar para si mesmo importam muito pouco, e assim por diante. Segue-se, então, que, ao aplicarmos o Princípio da Diferença - ao identificar os limites de sua aplicabilidade - nós devemos retornar à comparação informal das perdas que é central para a forma de contratualismo que eu descrevi.

²⁷ Por exemplo, Rawls, 1971, p. 141 e 148, embora essas passagens não possam distinguir claramente entre essa alternativa e uma suposição de benevolência.

²⁸ Rawls, 1971, p. 141.

V

Descrevi esta versão do contratualismo apenas em linhas gerais. Muito mais precisa ser dito para esclarecer suas noções centrais e elaborar suas implicações normativas. Espero ter dito o suficiente para indicar seu apelo como teoria filosófica da moralidade e como explicação da motivação moral. Apresentei o contratualismo como uma alternativa ao utilitarismo, mas o aspecto característico da doutrina pode ser destacado contrastando-o com uma visão um tanto diferente.

Às vezes se diz²⁹ que a moralidade é um dispositivo para nossa proteção mútua. De acordo com o contratualismo, essa visão é parcialmente verdadeira, mas incompleta de maneira importante. Nossa preocupação em proteger nossos interesses centrais terá um efeito importante sobre o que poderíamos razoavelmente concordar. Terá, portanto, um efeito importante sobre o conteúdo da moralidade se o contratualismo estiver correto. Na medida em que essa moralidade for observada, esses interesses ganharão com isso. Se não tivéssemos nenhum desejo de poder justificar nossas ações aos outros com fundamentos que eles pudessem razoavelmente aceitar, a esperança de obter essa proteção nos daria motivos para tentar inculcar esse desejo nos outros, talvez por meio de hipnose ou condicionamento em massa, mesmo que ele também significasse adquiri-lo para nós mesmos. Mas, dado que já temos esse desejo, nossa preocupação com a moralidade é menos instrumental. O contraste pode ser colocado da seguinte maneira. De um lado, a preocupação com a proteção é fundamental, e o acordo geral torna-se relevante como meio ou condição necessária para garantir essa proteção. Por outro lado, a visão contratualista, o desejo de proteção é um fator importante na determinação do conteúdo da moralidade porque determina o que pode ser razoavelmente acordado. Mas a ideia de acordo geral não surge como meio de garantir proteção. É, em um sentido mais fundamental, sobre o que a moralidade é.

Submissão: 20. 08. 2024

/

Aceite: 15. 09. 2024

²⁹ De diferentes formas, por G.J. Warnock, em Warnock, 1971, e por J.L. Mackie, em Mackie, 1977. Veja também as observações de Richard Brandt sobre a justificação, no Capítulo X de Brandt, 1979.