

A Teoria Ética de Hobbes: *uma exposição*

Elfio R. D. Mendes

*Dissertação apresentada como requisito parcial  
para a obtenção do grau de mestre em Filosofia,  
sob orientação do Prof. Dr. Nelson Boeira.*

Porto Alegre, setembro de 1996.



*Dedico esta dissertação aos meus pais, aos meus irmãos e a Rochele. Ao meu pai, por sua virtuosa tolerância; a minha mãe, por seu apoio intuitivo; ao Gustavo, pelo exemplo de entusiasmo intelectual; ao Sérgio, por seu inabalável otimismo. A Rochele, por seu amor.*





## SUMÁRIO

*Introdução* p. 5

*1. Arquitetônica da ética hobbesiana* p. 9

*1.1. Um modelo de teoria ética; 1.2. Uma definição de pessoa moral; 1.3. Motivação e conduta moral; 1.4. Uma forma de explicação de ações; 1.5. Um tipo de justificação de ações; 1.6. A relatividade temporal dos valores; 1.7. O desejo de autopreservação como causa original de toda ação; 1.8. A autopreservação como fim supremo; 1.9. A autopreservação como fim do "auto-interesse de longo prazo"; 1.10. O desejo de autopreservação visto como um desejo categórico; [1.11. Tipos e níveis de prova na teoria ética hobbesiana;] 1.12. Acordo moral e contrato social.*

alt.

*2. Teoria da motivação* p. 29

*2.1. O esquema conceitual da "psicologia materialista"; 2.2. Uma linguagem fisicalista para a psicologia; 2.3. O "materialismo metafísico"; 2.4. Construindo uma teoria da motivação com conceitos da "psicologia materialista"; 2.5. Construção teórica do desejo de autopreservação; 2.6. Um modelo biológico de explicação funcionalista; 2.7. Uma explicação funcionalista do mecanismo de autopreservação; 2.8. Limitações da explicação funcionalista; 2.9. Ações não podem ser explicadas pela "psicologia materialista"; 2.10. A intencionalidade das paixões.*

*3. Teoria do valor* p. 57

*3.1. A natureza dos juízos de valor; 3.2. A continuidade pessoal como maior bem; 3.3. O conceito de poder; 3.4. A cobiça expressa uma atitude valorativa em relação ao tempo; 3.5. O "egoísmo psicológico".*

*4. Teoria da racionalidade* p. 65

*4.1. A razão como cálculo; 4.2. Sobre a deliberação ou raciocínio prático; 4.3. Análise do conceito de utilidade; 4.4. A utilidade é uma medida de expectativa; 4.5. Utilidade e expectativas de poder; 4.6. Sobre a diversidade das concepções individuais de bem; 4.7. Juízos de valor subjetivos como "razões para agir"; 4.8. A concepção hobbesiana de razão prática; 4.9. A busca egoísta da felicidade; 4.10. Os papéis da razão e das paixões na deliberação; 4.11. As paixões e as falhas na escolha racional; 4.12. "Bem real" e escolha racional; 4.13. A prudência racional; [4.14. O objeto da "filosofia moral";] 4.15. A necessidade do acordo sobre um "valor comum"; 4.16. A paz como "valor comum"; 4.17. A concepção hobbesiana da moralidade.*

alt.

5. *Justificação das ações morais* p. 99

5.1. A “condição natural da humanidade”; 5.2. O que um “egoísta racional” busca?; 5.3. As “ações estratégicas” necessárias para a paz; 5.4. A natureza do “estado de guerra”; 5.5. O que significa: “algo é um meio necessário para produzir a paz”?; 5.6. A forma geral das “leis da natureza”; 5.7. O conceito de racionalidade moral; 5.8. O que há de errado com o egoísmo?; 5.9. O projeto de reconciliação da prudência com a moralidade; 5.10. Um acordo moral entre egoístas pode ser justificado?; 5.11. A racionalidade coletiva da “disposição para a paz”; 5.12. O “dilema do egoísmo”; 5.13. Como pode ser sempre racional agir de modo justo?; 5.14. Por que agir conforme as “leis da natureza”?; 5.15. A disposição de “abraçar a paz quando ela se mostrar viável”; 5.16. Sobre a obrigação de observar as “leis da natureza”; 5.17. Como é possível a cooperação entre egoístas?

*Apêndice I:*

*Existe uma dualidade de critérios de valor na ética de Hobbes?* p. 159

*Apêndice II:*

*Como é possível escapar do estado de natureza?* p. 162

*Apêndice III*

p. 170

*Bibliografia*

p. 172

## INTRODUÇÃO

O objeto de estudo desta dissertação é a teoria ética de Hobbes. Não tenho qualquer intenção de fazer uma apresentação completa sobre o tema; isto explica a ênfase no subtítulo: "uma exposição". Mais especificamente, trata-se de uma exposição delimitada: quero apenas desenvolver um "modelo" da teoria de Hobbes, quer dizer, num sentido não técnico, expor ordenadamente um modo de construção possível desta teoria. Por "construção" entendo, antes de tudo, o tipo de raciocínio <sup>he</sup> dual <sup>- r?</sup> que, para Hobbes, é a única via para se alcançar o conhecimento: derivar ("gerar" ou "construir" são os termos usados por Hobbes)<sup>1</sup> propriedades de um objeto a partir de elementos que descrevemos como partes de sua natureza ou, tomando como dadas tais propriedades, descobrir um modo de derivá-las (o que inclui descobrir de que elementos iniciais elas podem ser inferidas).

O raciocínio a que Hobbes se refere é aquele que os geômetras gregos denominaram "método da análise". Tal como apontou Pappus, importante matemático grego (sécs. II-III d. C.), existem, na verdade, dois tipos de análise: "a análise de 'problemas de prova', que visa estabelecer teoremas verdadeiros, e a análise de 'problemas de descoberta', que visa encontrar o que é desconhecido".<sup>2</sup> Os problemas 'de descoberta' e de 'prova' com os quais se ocupa a teoria ética hobbesiana têm, respectivamente, as seguintes formas:

i) *Dada a condição de guerra, encontre os elementos da natureza e interação humanas a partir dos quais uma tal condição é gerada.*

ii) *Dado que "os homens amam naturalmente a liberdade e o domínio sobre os outros" (não só por "força da avareza e da ambição", mas da radical situação de incerteza em que se encontram)<sup>3</sup> mostre que o "estado de guerra" é "uma consequência necessária das paixões naturais dos homens", a menos que, motivados pelo medo racional (só a razão pode dar lugar à "antevisão" de um "mal futuro") e querendo evitar sua própria destruição, os homens introduzam restrições ao seu comportamento de rapina.<sup>4</sup>*

<sup>1</sup> Ver, por exemplo, *Leviathan*, 46, 367.

<sup>2</sup> *Collectiones*, livro VII; conforme a tradução livre de G. Polya, *How to solve it*, p. 130.

<sup>3</sup> Cf., respectivamente, *Leviathan*, 17, 85; *De cive*, *Epístola dedicatória* e *Leviathan*, 13.

<sup>4</sup> Cf., respectivamente, *De cive*, *Epístola dedicatória*; *De cive*, I, II, última nota de Hobbes e *Leviathan*, 17, 85.

7 a resolução  
 i) não  
 contém  
 a resp.  
 lógico de  
 ii) não  
 modifica?

A prova dos teoremas da razão ou “leis da natureza” pode ser considerada um “problema auxiliar”<sup>5</sup> a ser enfrentado pela análise hobbesiana do “problema de prova”; quer dizer, um problema cuja solução serve como meio para mostrar a necessidade da consequência do “estado de guerra”.

O “modelo” aqui construído pressupõe a adoção da visão de uma teoria ética como um conjunto articulado de teorias componentes: da motivação, do valor, da racionalidade e da justificação das ações morais. Cada uma destas teorias poderia ser apresentada *sinteticamente*, isto é, como um conjunto de proposições deduzidas a partir de axiomas e postulados. A *Ética* de Spinoza é um exemplo deste modo de exposição; não é o caso, todavia, do projeto hobbesiano: Hobbes não pretende - presumo - apresentar sua teoria *more geometrico*, ele se movimenta no plano da *análise*.<sup>6</sup>

O que chamei de *arquitetônica* é, em parte, uma descrição bastante esquemática de alguns pontos de articulação lógica dessas teorias; mas é, principalmente, uma apresentação de pontos e distinções que parecem ser essenciais na construção de qualquer ética moderna pelo menos. Se isto for aceitável, tal modo de ver uma teoria ética pode ser proveitoso para entender o caso hobbesiano.

O ponto de partida da exposição analítica, balizada pelas observações gerais sobre a *arquitetônica*, é a *teoria da motivação* (capítulo 2). Para Hobbes, existe um motivo básico na psicologia de todo ser humano, que deve servir então de pressuposto de qualquer teoria sobre a natureza humana. Por isso, a ética, como teoria que ilumina alguns aspectos da natureza humana, é igualmente um ramo da psicologia (entendida com os olhos de um investigador do séc. XVII, isto é, algo como uma *metafísica da motivação humana*). Este motivo é o *desejo de autopreservação*. Pretendo oferecer uma interpretação do papel deste desejo cuja associação com o uso deliberativo da razão é capaz de esclarecer uma afirmação fundamental para a construção da teoria hobbesiana:

<sup>5</sup> Uma definição deste termo dentro da *heurística* (*arte de “resolver” - inventar soluções para - problemas*) pode ser encontrada no livro de Polya, *op. cit.*

<sup>6</sup> Descartes, nas *Respostas às segundas objeções*, contrasta o “método de síntese” - em que cada passo na cadeia de consequências está contido nos passos antecedentes - com o “método da análise”, que “mostra o verdadeiro caminho pelo qual uma coisa foi metodicamente descoberta e revela como os efeitos dependem das causas”. Descartes acrescenta a seguir: a síntese é o método mais apropriado para as demonstrações formais em geometria; a análise, que não deixa de ser uma “espécie de demonstração”, é mais apropriada “às matérias que pertencem à Metafísica”, bem como ao seu ensino. Cf. pp. 166-167.

"Assim cheguei a duas máximas da natureza humana - uma que provém de sua parte concupiscente, que deseja apropriar-se do uso daquelas coisas nas quais todos os outros têm igual participação, outra, procedendo da parte racional, que ensina todo homem a fugir de uma dissolução antinatural, como sendo este o maior dano que possa ocorrer à natureza."<sup>7</sup>

A seguir, apresento a teoria hobbesiana do valor - derivada da sua teoria da motivação -, cujo núcleo expõe uma condição de possibilidade do acordo racional entre egoístas para a instituição da moralidade. Porém, tal acordo não pode ser completamente entendido se não se explicar o papel da razão na determinação do acordo mínimo entre as diferentes concepções privadas de bem: todos os homens capazes de deliberar racionalmente estão de acordo sobre o valor da paz, quer dizer, concordam que ela deve ser buscada porque é um meio necessário para a realização de qualquer plano de vida racional. No "estado de guerra de todos contra todos", estes planos de vida são frustrados, impossibilitando a realização das concepções individuais de bem. Neste sentido, o objeto desse acordo racional são as regras da moralidade pública, que expressam condições mínimas para a existência da vida moral individual em sociedade.

Assim, o capítulo 4 apresenta a teoria da racionalidade pressuposta na teoria ética contratualista de Hobbes. Em primeiro plano, é analisada a concepção de deliberação; o papel das paixões como elementos explicativos centrais das chamadas "falhas da decisão racional" também é aí exposto. São estas "falhas" - predominantemente o tipo explicado pela preponderância da disposição de agir de modo a ter, no presente, sempre mais e mais: a ganância, avidez ou cobiça (o termo hobbesiano é "greed") - que explicam por que a "condição natural da humanidade" é um estado de guerra.

Por fim, no capítulo 5, é apresentada a teoria da justificação das ações morais, isto é, o argumento de Hobbes para mostrar que, dada a natureza humana, as ações prescritas pelas "leis da natureza" são os melhores meios para a produção da paz. "A paz" refere-se aqui a uma rede de interações humanas sob a qual a vida de cada indivíduo deixa de estar sob ameaça: sob certas restrições, então, cada um é livre para buscar sua própria concepção de bem. Spinoza sintetizou o que pode ser tomado como o resultado principal - e aparentemente paradoxal - dessa parte da teoria hobbesiana: "o homem que é dirigido

<sup>7</sup> De cive, epístola dedicatória; trad. de Renato Janine Ribeiro.

leis da natureza humana ?

pela razão é mais livre no estado onde vive segundo o decreto comum, do que na solidão onde só obedece a si próprio”.<sup>8</sup>

Já os apêndices I e II apresentam exercícios de argumentação visando resolver de modo um pouco mais rigoroso dois problemas bastante áridos que surgem na construção da teoria ética de Hobbes. O apêndice I tenta mostrar que uma suposta incoerência na teoria do valor hobbesiana é apenas aparente. O apêndice II expõe de modo mais minucioso o problema enunciado pela questão “como é possível a cooperação entre egoístas?”, cuja resposta está apoiada, em última análise, em toda a construção filosófica hobbesiana da natureza da moralidade. Da existência de uma resposta plausível parece depender a força desta construção teórica. Ainda que de modo não definitivo, os apêndices devem servir para expor, de modo um pouco mais refinado, parte da complexidade da estrutura teórica em estudo. O apêndice III apresenta um esclarecimento e uma justificação de afirmações feitas de modo excessivamente dogmático na seção 2.7.

O objetivo mais geral deste trabalho - que, infelizmente, deixa muitos problemas importantes apenas assinalados - é mostrar que a teoria ética de Hobbes pode ser considerada como uma alternativa razoável (não necessariamente a melhor) na construção de uma explicação (não necessariamente a mais abrangente) da natureza da moralidade.

---

<sup>8</sup> *Ética*, quarta parte, proposição LXXIII.

## I. ARQUITETÓNICA DA ÉTICA HOBBSIANA

"No primeiro período, todos os moralistas ortodoxos ocupavam-se em refutar o Hobbesianismo. Mas este sistema, embora baseado no materialismo e no egoísmo, também foi sustentado como eticamente construtivo"

Henry Sidgwick, *The methods of ethics*

### 1.1. Um modelo de teoria ética.

Segundo uma certa tradição da filosofia moderna,<sup>9</sup> uma *teoria ética* é formada por um conjunto ordenado de proposições gerais acerca dos constituintes da chamada *conduta moral*. Desde um ponto de vista teórico, dizemos, de modo genérico, que alguém se conduz moralmente se as suas ações podem ser descritas e, também com base nos princípios de alguma teoria ética, justificadas como moralmente boas e corretas. Indicar a base (os princípios) e o procedimento, a partir dela, de justificação moral das ações é uma das funções normativas de uma teoria ética.<sup>10</sup>

É igualmente admitido que, como parte da explicação das ações morais, a teoria deve indicar quais são os *motivos* que governam a conduta moral, isto é, aqueles motivos que garantem que as exigências estabelecidas pelos *princípios éticos* tenham força e, conseqüentemente, efetivamente causem as ações. Temos, desta maneira, um grupo mais restrito de proposições cujo objeto são os motivos considerados como genuinamente morais. Por isso, parte do papel de uma teoria ética deve ser realizada por uma *teoria da motivação*.

Numa teoria ética, algumas proposições fazem parte de um esquema construído visando oferecer uma *fundamentação* dos princípios. A fundamentação é feita por um argumento por meio do qual se busca responder à pergunta: por que devemos agir obedecendo a princípios éticos em geral (e aos de tal teoria em particular)? Quando

<sup>9</sup> Ver, por exemplo, o tipo de estrutura que Bernard Williams denomina como *teoria ética*; *Ethics and the limits of philosophy*, principalmente p. 72.

<sup>10</sup> Vale lembrar aqui a definição de Kant no início de *Em torno ao tópicó*: "Talvez isso seja correto em teoria, mas não serve para a prática": "Denomina-se como *teoria* inclusive um conjunto de regras práticas, sempre que tais regras sejam pensadas como princípios, com certa universalidade".

do ponto de vista teórico?

perguntamos pela *validade* de princípios éticos, precisamos ter em mente este tipo de discussão.

Um processo de *validação* ou *fundamentação* é concebido como envolvendo o recurso a algo que seja irrecusável por parte de seres racionais. No caso da ética, concebe-se as exigências feitas sobre nossas ações de forma a não podermos, enquanto seres racionais, delas escapar; caberia a nós, portanto, somente a *obediência*.<sup>11</sup> Assim, uma investigação sobre a fundamentação da ética deve apontar as condições que tornam os imperativos morais (sejam eles quais forem) fonte de uma conduta cuja omissão não pode ser racionalmente justificada, nem explicada. O que se busca é oferecer uma argumentação que seja uma prova da "necessidade racional de se agir eticamente"<sup>12</sup> ou, dito de outro modo, "uma justificação para ser moral".<sup>13</sup> "Necessidade racional" indica aqui duas condições necessárias (nenhuma delas por si só suficiente, pois são logicamente independentes - o que é devido, em última análise, à impossibilidade lógica de derivar-se um "deve" de um "é"):

a) existe uma razão para *afirmar* ou *sustentar* a relação - entre uma ação (ou sua omissão) e um estado de coisas que depende, de algum modo, dela - descrita pelos imperativos morais;

b) existe uma razão para *aceitar* os imperativos morais como *prescrições*, ou seja, uma razão para *obrigar-se* a fazer o que eles prescrevem.<sup>14</sup>

Na condição expressa em *a*, o termo *depende* pode ser assim interpretado: um estado de coisas qualquer depende de uma ação na medida em que é produzido <sup>como</sup> um dos seus resultados ou *conseqüências*. Segundo as teorias éticas deontológicas - a kantiana, por exemplo -, uma ação genuinamente moral não pode ser justificada pelos seus resultados ou *conseqüências*, sejam eles quais forem; as razões para uma tal ação residem, em última

<sup>11</sup> Warnock sublinha a dificuldade envolvida no problema de *fundamentação* no caso da ética: "Gostaria agora de me perguntar se existe algum sentido, e neste caso qual, em que um ser racional se encontre obrigado - enquanto racional - a 'aceitar' princípios morais ou a reconhecê-los e pô-los em prática através de seus juízos e suas ações... Poderia um ser semelhante rechaçar a moralidade sem perda de sua racionalidade? *The object of morality*, p. 152.

<sup>12</sup> Peter Singer, *Practical ethics*, p.207.

<sup>13</sup> Thomas Nagel, *The possibility of altruism*, p. 3; grifo do autor.

<sup>14</sup> Não parece haver discordância sobre o fato de que imperativos morais são *normas* ou *prescrições*. Ora, parece ser possível para alguém reconhecer a existência de uma norma - bem como não discordar do fato de que, quando obedecida, "ela produz" tais e tais resultados - e, ao mesmo tempo, não se considerar obrigado a fazer o que nela está prescrito. A distinção conceitual garantida por tal possibilidade está na base da feita entre as razões referidas nas condições *a* e *b*.



análise, no fato de ser racional para alguém *querer agir* do modo *correto*, determinado por traços que se aplicam a quaisquer ações morais, independentemente de quais possam ser as suas conseqüências.<sup>15</sup>

Segundo as teorias éticas conseqüencialistas - o que é o caso da hobbesiana -, uma ação moral deve ser justificada pelas suas conseqüências.<sup>16</sup> Basicamente, as razões para aceitar as ações prescritas pelos imperativos morais residem no fato de tais conseqüências serem aceitáveis como boas. Mas boas para que ou segundo qual padrão? Esta é uma pergunta que a teoria do valor de Hobbes deve poder responder.

Uma prova da condição expressa em *b* consiste na busca de um "ponto arquimediano" que tenha *força motivacional* para levar alguém a orientar a sua conduta pelo que é recomendado por imperativos morais. Uma tal prova não consiste em tão

<sup>15</sup> Para ser mais preciso: "conseqüência" pode ser, mesmo neste contexto, um termo enganador. Como observou Ted Honderich, para qualquer teoria moral conseqüencialista, as *razões para agir* possíveis servem para avaliar a satisfação relativa a dois tipos de estados de coisas:

i) os que se *seguem* do fato de alguém fazer *x* (que costumamos chamar de "conseqüências" da ação *x*);

ii) os que *consistem* no próprio fato de alguém fazer *x* (que podemos chamar de "conseqüências intrínsecas" da ação *x*).

Cito esse autor: "uma ação pode assim ser a coisa certa ... em grande parte devido à satisfação ou enriquecimento que tem a pessoa *em fazê-la*". *Introdução* ao artigo *A critique of utilitarianism*, de Bernard Williams. In Honderich, T. & Burnyeat, M. (eds.). *Philosophy as it is*, p. 33. Depois de definir "conseqüência" de modo tão amplo, parecendo ter se apropriado de "todas as possíveis razões para agir que podem existir", que razões restam para embasar as avaliações feitas por uma *moralidade não conseqüencialista*? Uma tal avaliação teria a seguinte forma: um estado de coisas é melhor que suas alternativas porque "consiste em se estar fazendo a ação correta". *Idem*. [Acrescento: isto é que se chamaria de uma *razão objetiva* para agir moralmente. Resta tudo a esclarecer sobre natureza do *fato* (psicológico? fisiológico?) que poderia garantir - sem cairmos numa explicação circular - a correção de uma ação.]

<sup>16</sup> Os termos mais comuns da oposição genérica aqui apresentada são "teorias deontológicas" e "teorias teleológicas". Anscombe considera que os "conseqüencialistas" compartilham a concepção de que a ação moralmente correta é "a ação que produz as melhores conseqüências possíveis". *Modern moral philosophy*, p. 9. Segundo Nancy Davis, "para os teleologistas, a correção ou não das nossas ações é determinada por um cálculo comparativo das suas conseqüências". *Contemporary deontology*, em Singer (ed.) *A companion to ethics*, p. 205.

O uso de um ou outro destes termos apenas ajuda a destacar, de diferentes ângulos, um mesmo aspecto da estrutura lógica destas teorias. Assim, se nomeamos uma teoria ética como "teleológica", estamos enfatizando o fato de que nela o que é "correto" fazer é definido como o que *maximiza* a produção de um bem definido independentemente. Ao contrário, quando dizemos que uma teoria é "conseqüencialista", estamos enfatizando o tipo de conexão que existe entre a ação moralmente correta e o que é avaliado como moralmente bom: aquela ação é a que produz os *melhores* resultados, que são assim avaliados por um critério independente do que seja considerado como correto fazer. Ver a discussão já clássica de John Rawls sobre a distinção "teleológico *versus* deontológico" aplicada às "estruturas" de teorias éticas em *A theory of justice*, pp. 24-25 e 30.

somente exibir uma razão para a ação moral, mas em exibir uma conexão entre tal razão e um motivo pelo qual as pessoas são suscetíveis de influenciar-se. Exibir uma tal conexão consiste basicamente em mostrar que é característico destas pessoas ter um motivo que explica por que uma razão para agir moralmente é capaz de fazê-las *querer* assim agir.

A condição expressa em *b* é portanto a que de algum modo garante que o juízo indicado em *a* possa adquirir força, isto é, adquirir o poder de "colocar-nos em movimento" ou "ter influência sobre nossas ações".<sup>17</sup> Esta operação de atribuição de força prescritiva, explicável por uma caracterização adequada da motivação, transforma o juízo "Devemos fazer *x*" em outro juízo, do tipo "Queremos fazer *x*" (expressão de um "ato da vontade").<sup>18</sup>

### 1.2. Uma definição de pessoa moral.

Apresento uma definição, por assim dizer, "mínima" de *pessoa moral*: "Pessoas morais ... são aquelas que podem racionalmente seguir princípios morais mesmo com um custo (líquido) para os seus valores; elas agem moralmente porque têm boas razões para assim fazer, quer dizer, é racional para elas assim fazer".<sup>19</sup> Embora seja uma definição algo vaga (por conter termos interpretáveis diferentemente), ela parece ser *neutra*, no sentido de poder revelar um nível sobre o qual diferentes autores estariam de acordo.

Para efetuar uma análise deste conceito, é possível começar orientando-se por um método de molde kantiano. Seguindo <sup>o</sup> ~~um tal método~~, pergunta-se pelas condições que tornam possível a conduta moral, tal como aparece implicada na definição acima. Isto significa indagar, entre outras coisas, pelos seguintes pressupostos do conceito e da prática ao qual ele se aplica:

- i) quais são os valores referidos e de que modo eles operam na conduta moral?
- ii) o que significa dizer que "alguém tem boas razões para fazer *x*", ou, mais abstratamente, que "é racional para tal pessoa fazer *x*"?
- iii) de que modo valores e razões estão relacionados?

<sup>17</sup> Cf. Anscombe, *Modern moral philosophy*, p.7.

<sup>18</sup> Podemos dizer que este resultado pressupõe uma condição forte: que a pessoa para a qual a prova é apresentada está interessada em agir racionalmente. Dito de outra forma, um argumento para "justificar a moralidade" só têm relevância para aquelas pessoas que estão de antemão dispostas a ouvi-lo precisamente como uma *justificação racional*.

<sup>19</sup> Gaus, *Value and justification*, p. 17. [ALRES.]

iv) que tipo de agente (crenças, interesses, capacidades, desejos, motivações, etc.) é capaz de ponderar valores e razões tal como é pressuposto de "pessoas morais"? E de que modo é feita uma tal ponderação?

Estas indagações - que são, mais precisamente, objeto de estudo da chamada "psicologia moral" - servirão de fio condutor para boa parte da discussão que se segue neste e nos demais capítulos.

### 1.3. Motivação e conduta moral.

Para Hobbes, entre as condições que tornam possível a conduta moral, existe um motivo fundamental: o desejo de autopreservação. Para esclarecer o papel explicativo deste desejo na teoria da motivação, é preciso introduzir alguns conceitos básicos. Primeiramente, o que é um *motivo*? De modo esquemático, há duas formas de descrever o significado deste termo. Um motivo é tudo o que é referido pelas expressões que podem aparecer completando sentenças da forma "Fulano agiu por causa de ...". Sob outra descrição, um motivo é a causa (origem) do movimento de uma ação, a sua fonte psíquica de energia ou, tal como Kant o denominou, o seu *incentivo* (*Triebfeder*).<sup>20</sup>

No entanto, esta tentativa de definir o termo não é por si mesma muito esclarecedora. Isto porque as duas descrições não são co-extensionais. Por exemplo, o primeiro tipo de sentença pode ser completado por uma crença ("João matou sua esposa por causa de sua suspeita de ser ela uma adúltera"), e crenças, supõe-se, não podem por si só causar uma ação. A ambigüidade da expressão "por causa de ..." é geradora deste problema: pode referir-se, entre outras coisas, a "por acreditar que ...", "por visar ...", "por ser movido (levado) por ...".

Para se poder atribuir a uma pessoa *A* um motivo para fazer uma ação *x*, são necessárias duas condições:

- i) *A* crê que *x* é um meio para realizar um fim *y*;
- ii) *A* quer (i. e., deseja) realizar *y*.<sup>21</sup>

<sup>20</sup> Cf. R. Bittner, *What reason demands*, p. 7.

<sup>21</sup> A justificação para a inclusão destas notas características na definição do conceito de *motivo* é a seguinte: ela é a teoricamente mais produtiva quando se quer construir explicações que dêem conta do caráter essencialmente intencional do que denominamos propriamente de *ações*. Além disso,

(o conjunto das coisas que *A* quer realizar, dos seus *fins* ou objetivos, é o conjunto dos valores de *A*)

A condição *ii* indica que não pode haver motivo para agir que não esteja baseado num desejo. Deve ser possível mostrar a conexão entre este desejo e o desejo de fazer *x*; esta é uma das tarefas da teoria da deliberação: mostrar como deliberamos sobre meios para fins desejados. Acrescento uma observação importante: para *A* ter um motivo para fazer *x* não é necessário que *A* queira fazer *x*, porque é possível que seu motivo para não fazer *x* seja mais urgente. Visto de outro modo: não é possível deduzir "*A* quer fazer *x*" das proposições *i* e *ii*. Penso que se isso não fosse assim, não poderia haver escolha entre fazer ou não fazer *x*, porque não poderia haver a deliberação<sup>22</sup> que lhe deve preceder logicamente e da qual deve ser possível resultar o desejo de fazer *x* ou o de deixar de fazer *x*.

#### 1.4. Uma forma de explicação de ações.

Pode-se explicar uma ação indicando quais os motivos que levam alguém a fazer o que faz. Neste tipo de explicação, emprega-se uma noção explicativa (não-normativa) de *racionalidade*. Explicamos e interpretamos os movimentos realizados por alguém como uma ação racional porque entendemos, (num sentido básico,) o seu comportamento como expressando uma coerência mínima entre suas crenças, desejos e valores ou preferências. O tipo de explicação assim formada é o que se chama de "explicação teleológica ou fundada em razões". É distinta das explicações das ciências naturais, onde os fenômenos são relacionados causalmente por meio de leis. Nestas ciências, concebe-se os fenômenos como gerados por forças cujo "mecanismo" de operação pode ser descrito como "causalidade cega".

---

também facilita a introdução de outros conceitos-chave, como é o caso dos envolvidos na distinção entre *imperativos categóricos* e *hipotéticos*.

Um *motivo* pode ser definido pelos conceitos de *fim* e *meio*. "(...) *fim* é o motivo direto de um ato da vontade; *meio*, o motivo indireto... Ser *fim* significa: ser querido". Schopenhauer, *Über das Fundament der Moral*; p. 71 na trad. brasileira.

<sup>22</sup> Como Aristóteles assinala, nós só deliberamos sobre meios possíveis para fins desejados; e isso pressupõe poder *escolher* entre meios alternativos conducentes a estes fins.

Muitos filósofos fizeram referência ao que Davidson chama de "explicações fundadas em razões". Este autor lembra uma passagem onde David Hume expressa "com energia" a estrutura (*pattern*) deste tipo de explicação:

*"Pergunte a um homem por que faz exercício. Responderá: porque deseja manter-se sã. Se então você lhe pergunta por que deseja a saúde, ele rapidamente responderá que a deseja porque estar doente é penoso."*<sup>23</sup>

Segundo Davidson, esta estrutura é familiar a ponto de nem mesmo nos permitir perceber a sua sutileza. Como se trata de explicar ações, a explicação estabelece uma relação entre dois fatores: entre um valor, uma meta, uma preferência ou um desejo (*want*) do agente, e uma crença de que, ao fazer o que pretendemos explicar, o agente pode promover um valor ou meta relevante, ou bem que age de acordo com sua preferência ou desejo. Dados estes fatores, a relação entre eles deve satisfazer duas condições para estruturar uma "explicação fundada em razões":

*"A ação, por uma parte, e o par crença-desejo que proporciona a razão, por outra, têm que estar relacionados de dois modos diferentes para que se produza uma explicação. (Primeiro) tem que haver uma relação lógica. As crenças e os desejos têm um conteúdo, e estes conteúdos têm que ser tais que impliquem que há algo de valioso ou desejável na ação. Assim, o homem que vê algo valioso na saúde e que o exercício fará dele um homem saudável, pode concluir que há algo desejável no exercício; o que pode explicar por que faz exercício. (Em segundo lugar) as razões que um agente tem para atuar, se têm de servir para explicar a ação, tem que ser as razões pelas quais agiu; as razões têm que ter desempenhado um papel causal na realização da ação [i. e., as crenças e os desejos são causas das ações para as quais são razões]".*<sup>24</sup>

<sup>23</sup> Citado em *Las paradojas de la irracionalidad*, p. 5. Esta passagem, do apêndice 1 de *An inquiry concerning the principles of morals*, é parte do argumento que Hume monta para concluir que, a menos que concebamos como possível um "progresso *ad infinitum*", a cadeia de razões que podem ser oferecidas para explicar por que desejamos as coisas que desejamos deve ter um fim. Se isto é assim, então deve haver algo desejável por si mesmo, um "fim último", uma coisa que não pode ser razão para nenhuma outra, algo que é desejado não por ser meio para alcançar uma recompensa associada a um fim, mas - como Hume pensa ser o caso da virtude - "meramente pela satisfação imediata que produz". *Op. cit.*; pp. 161-162 na ed. argentina.

<sup>24</sup> *Idem*. Grifo meu. Algumas vezes, explicar ações envolve descobrir motivações pouco aparentes, e mesmo *inconscientes*; isto é uma das doutrinas centrais da teoria psicanalítica desenvolvida por Freud. Esta teoria ampliou o alcance da explicação fundada em razões, pois ampliou a variedade dos fenômenos que se "explicam racionalizando": acontece que temos razões para nossos esquecimentos, medos exagerados, deslizes verbais ou gestuais, comportamentos ritualísticos, etc., ainda que possam gerar ações ou atitudes que, levando-se em conta tudo o que é relevante (*everything considered*), não se considere como as melhores. Aqui reside, segundo Davidson, um dos "paradoxos da irracionalidade": "todas as ações intencionais, sejam ou não irracionais em algum sentido, contêm, em seu núcleo, um elemento racional". *Ibid.*, p. 5. A irracionalidade de que fala

exercício-saúde

razões | explicar | causas

A virtude  
é a  
sua própria  
recompensa

### 1.5. Um tipo de justificação de ações.

Explicar uma ação por meio deste procedimento não é suficiente para justificá-la, porque indicar os motivos para *A* fazer *x* não equivale a indicar as melhores razões para *A* fazer *x*, nem a dizer que *A* tem boas razões para fazer *x*, nem muito menos, que tais motivos são suas únicas razões para fazer *x*. Se nos limitarmos à concepção de "razões relativas a desejos",<sup>25</sup> dizer que *A* tem uma boa razão para fazer uma ação *x* equivale a dizer que a crença de *A* de que *x* é um meio para um fim desejado por *A* é uma crença justificada, ou ainda, que *A* tem razão para assim crer.

Por sua vez, justificar uma crença consiste essencialmente em apresentar boas evidências para a sua verdade. Buscar oferecer razões para crenças que apoiam motivos para agir de determinado modo e não de outro qualquer é o objetivo de quem está interessado em justificar sua ação e não somente em descrevê-la (quer dizer, não somente em torná-la inteligível). Para tanto, é necessário pressupor uma regra ou padrão de racionalidade para orientar a sua decisão. Saber qual é o alcance ou limite da razão para orientar escolhas é tarefa de uma *teoria da racionalidade*.

Saber se a razão é suficiente para orientar escolhas de ações morais também é tarefa desta teoria se a concebemos como um componente de uma teoria ética. Para Hobbes, pelo menos, existe uma motivação não-racional para agir moralmente. Isto, como veremos, quer apenas dizer que boa parte do "acerto moral" da deliberação ou raciocínio prático é avaliado pelo "acerto" da motivação ("acertar" a motivação é uma operação que interfere na própria aceitação da adequação instrumental entre meios e fins, basicamente porque interfere no que se considera como evidência para esta adequação).

---

Davidson está basicamente associada a *razões* para crer, desejar, agir, *que não são as melhores* quando comparadas a outras pertencentes ao que Davidson chama de "estrutura de razões" de um agente, na qual estão entremeados crenças, expectativas, suposições, preferências, desejos. *Ibid.*, pp. 11-13.

<sup>25</sup> Esta pressuposição é contrária à tese de que "razões para agir são logicamente independentes de [quaisquer] desejos". Para uma defesa desta tese, ver G.R. Grice, *Motive and reason* in J. Raz (ed.), *Practical reasoning*, principalmente pp. 168-170. Grice apoia sua tese na pressuposição de que algo pode ser do interesse de alguém fazer independentemente de quaisquer que sejam seus desejos. Esta é uma pressuposição, veremos, incompatível com a teoria do valor hobbesiana; por isso, se meu objetivo é apresentar uma interpretação possível da estrutura teórica da ética de Hobbes, devo deixá-la de lado.

*acerto da motivação*

Se pressupomos a prudência como tal padrão, podemos aceitar um tipo de proposição cuja verdade é condição suficiente para haver uma boa razão para fazer  $x$ . É a proposição "Fazer  $x$  é do auto-interesse de  $A$ ", que deve ser analisada como " $x$  é, nas circunstâncias dadas, o único (ou o melhor) meio disponível para  $A$  realizar seu auto-interesse" ou "É racional para  $A$ , nas circunstâncias dadas, fazer  $x$ ". Devemos aqui supor que  $A$  quer realizar tal fim e que este desejo pode ser representado por uma preferência mais forte que as demais.<sup>26</sup> Com estas suposições, evitamos postular a existência de valores objetivos ou fins intrinsecamente desejáveis, concebidos como representando boas razões para alguém fazer uma ação  $x$  independentemente de qualquer referência (implícita ou explícita) ao que tal pessoa deseja ou a seus objetivos e propósitos (sejam eles auto-interessados ou não).<sup>27</sup>

Resumindo o ponto mais relevante: a prudência - cuja análise por meio do conceito de "auto-interesse" será desenvolvida adiante - é uma base racional para avaliar que, se tenho boas evidências para crer que uma ação resulta na satisfação daquilo que prefiro ao realizá-la, então tenho boas razões para praticá-la; e se tenho evidências do contrário, quer dizer, que ela não resulta em tal satisfação, então tenho boas razões para deixar de praticá-la.

Resta elucidar o seguinte: dizer que proposições do tipo "Fazer  $x$  é do interesse de  $A$ " funcionam como boas razões para  $A$  fazer  $x$  não equivale a dizer que todo e qualquer fim ou desejo de  $A$  é uma boa razão prudencial para  $A$  fazer  $x$ . Se, como veremos,

<sup>26</sup> Ganha-se ao empregar o termo "*preferência*" quando se trata de analisar deliberações e escolhas em que há conflitos entre desejos, interesses, ou ainda, carências. Preferências, genericamente, podem ser tomadas como desejos de 2ª ordem, i. e., desejos que servem para hierarquizar outros desejos; e isto também vale para interesses e carências. Uma *carência* é sempre entendida como uma qualificação de um desejo cuja satisfação é urgente (o grau dessa *urgência* ou *necessidade* expressa o lugar que o objeto do desejo - o *fim* - ocupa numa hierarquia de preferências). O que cada pessoa escolhe, quando diante de diferentes alternativas de ação, depende, na ausência de coerção, de suas preferências. A racionalidade de uma escolha pode então ser observada num exame da coerência de suas preferências. A isto se presta, por exemplo, o caso do capitão do navio que, surpreendido por uma tremenda tempestade, *precisa escolher* entre salvar a carga que transporta e salvar sua vida e a da tripulação (com diferença em alguns detalhes, este é o mesmo exemplo dado por Aristóteles na *Ética a Nicômaco*, III, I).

<sup>27</sup> Por outro lado, para que no caso a ação  $x$  seja do interesse de  $A$  fazer (o que equivale a dizer: para que  $A$  tenha a melhor razão para fazer  $x$ ), não é necessário que  $A$  queira fazer  $x$ . A tarefa de explicar como é possível que as pessoas deixem, algumas vezes, de reconhecer que fazer uma determinada ação é do seu interesse deve ser realizada por uma *teoria do erro da escolha racional*. Ver 4.11 e 4.12.

entendermos *o interesse que define a prudência* como um valor de 2<sup>a</sup> ordem, como função de um desejo de 2<sup>a</sup> ordem, então a busca da satisfação de alguns desejos não pode ser justificada prudencialmente.

### 1.6. A relatividade temporal dos valores.

O valor que se dá a cada consequência, quando se avalia ações alternativas, também depende da maior ou menor importância dada à sua maior ou menor proximidade do presente, isto é, do momento em que a ação é escolhida. E, mais importante para os objetivos deste trabalho, pelo menos na caracterização hobbesiana completa do "estado de natureza", deve-se incluir o fator *tempo*. Quer dizer, explicando-o melhor, descrevemos tal estado como aquele em que é essencial o fato de que "as pessoas vivem no presente e se preocupam apenas consigo mesmas".<sup>28</sup> Há uma forte razão para incluir este fator como essencial: nas escolhas que cada pessoa faz, ela atribui um *valor presente* às consequências que motivam a sua ação. Isto é, a cada consequência prevista *agora*, a pessoa associa *agora* um valor (i. e., atribui-lhe *um valor presente*); e mesmo quando alguém escolhe fazer *no futuro* alguma coisa, a sua intenção no momento em que escolhe é fazer *mais tarde* o que *agora* prefere: tal pessoa pretende *agora* fazer *então* o que prefere *agora*.<sup>29</sup>

Contudo, esta razão não dá conta inteiramente do significado da expressão "as pessoas vivem no presente". Esta última também descreve um traço próprio daqueles que se encontram no estado de natureza: quanto mais distante o *tempo futuro* de um dado resultado da ação, tanto menor o *valor presente* que lhe é atribuído. Tal valoração não tem necessariamente a mesma origem motivacional: nem todos que buscam uma gratificação de curto prazo o fazem movidos pela *cobiça*. Num mundo povoado por "predadores"<sup>30</sup>, por

<sup>28</sup> Jon Elster, *Nuts and Bolts for the social sciences*; p. 60 ed. bras.

<sup>29</sup> Cf. Elster, *op. cit.*, pp. 64-65 ed. bras.

<sup>30</sup> O termo "predador" é usado na teoria dos jogos para designar o tipo de jogador cujo comportamento é análogo ao dos predadores na teoria biológica sobre as chamadas "cadeias de alimentação". Ele pode ser feito significar o que Hobbes denominou de "*competitor*": "a competição (*competition*) pela riqueza, a honra, o mando e outros poderes leva à luta, à inimizade e à guerra, porque o caminho seguido pelo competidor (*competitor*) para realizar seu desejo consiste em matar, subjugar, suplantar ou repelir o outro". *Leviathan*, 11, 47-48. (Todas as citações desta obra, salvo algumas correções indicadas, são da tradução brasileira da ed. "Os Pensadores"; os números indicam



exemplo, não é razoável deixar de "descontar o futuro", esperando resultados cujas recompensas, ainda que maiores, são proteladas.

O comportamento dos "predadores" gera assim um "efeito dominó": cada um se fixa no que pode garantir imediatamente, simplesmente porque protelar suas recompensas pouco serve para garantir a sua sobrevivência. Este é o comportamento de um agente cujo tipo de psicologia leva-o a dar mais valor às conseqüências mais próximas de uma ação simplesmente pela perspectiva de vir a desfrutá-las antes. Mas também é o comportamento de um outro agente, que sabendo da existência de uma tal psicologia, dá mais valor às recompensas possíveis de serem obtidas imediatamente, por ter prudencialmente passado a supor que, em cada interação, seu oponente é movido pela cobiça.

### 1.7. O desejo de autopreservação como causa original de toda ação.

( Segundo a concepção de Hobbes, a causa original de toda ação é sempre um desejo. Isto porque somente desejos têm força motivacional, isto é, podem dar origem ao movimento de uma ação. Há, no entanto, diferentes processos de formação de desejos. Há desejos que simplesmente nos assaltam, analogamente àquelas crenças que temos pelo simples fato de percebermos alguma coisa. Um exemplo disso é o fato de acontecer a alguém sentir fome por falta de comida.

A fome, por sua vez, pode levar tal pessoa a desejar sair para comprar algo, quando vir a constatar que a sua geladeira está vazia. Neste caso, podemos dizer que o desejo de sair para comprar algo foi de algum modo gerado pela fome. Mas esta passagem não se deu diretamente, do mesmo modo que não se pode concluir que  $x$  é um elefante somente a partir da crença de que  $x$  é um animal. Aquele desejo resultou de um processo de deliberação, no término do qual a pessoa poderia ter decidido (por ter vindo a desejar) ir a um restaurante, à casa de um amigo, etc. Outros fatores (crenças, informações, preferências, valores, etc.) interferiram igualmente na sua formação. É importante notar

---

respectivamente o capítulo e as páginas na ed. original de 1651.) No latim clássico, *praedator* é usado para designar todo aquele que busca violentamente fazer outro de presa (*praeda*); vale dizê-lo de um saqueador, um pirata, um caçador e, por extensão, quem, com avidez e violência, busca conseguir o que deseja.

que, desejos imotivados podem funcionar como *impulsos imediatos* para ações, e por isso mesmo, também contrariar desejos motivados.

Usando da distinção estabelecida pelo exemplo acima, podemos dividir os desejos em duas grandes categorias: desejos imotivados e desejos motivados (i. e., formados como meios de satisfação de outros desejos). Tendo introduzido esta distinção, pode-se especificar a tese de Hobbes: a causa original de toda a ação é sempre um desejo imotivado - um desejo que não é o de um meio para satisfação de (o prazer de realizar) outro desejo -, ainda que nenhuma ação possa derivar diretamente dele, não só por haver alterações produzidas colateralmente às modificações na cadeia de razões que leva à escolha de uma ação, mas porque, tal como Hume ensinou, somente desejos de objetos particulares podem produzir inclinações direcionadas (que “focalizam”, por assim dizer, um objeto como meta) para a satisfação.

*“Em todos estes casos há uma paixão que aponta imediatamente ao objeto e que o constitui em nosso bem ou felicidade, assim como há outras paixões secundárias que logo surgem e o perseguem como parte de nossa felicidade... Se não houvesse apetite de nenhum tipo, antecedente ao amor a si mesmo, esta inclinação dificilmente poderia ser exercida alguma vez.”<sup>31</sup>*

A autopreservação não é conteúdo de nenhum desejo particular, mas é afirmada como uma hipótese fundamental sobre a natureza humana: a existência de um desejo que serve de fonte motivacional última de toda ação, ainda que, em certos casos, esta ação seja incorretamente avaliada como um meio de satisfação deste desejo (um aspecto a esclarecer: de que tipo de *erro* estamos falando?).

No capítulo 4 será apresentada mais detalhadamente a concepção hobbesiana de deliberação, que está sendo aqui apenas vagamente pressuposta. Ainda assim, é possível antecipar um ponto importante relativo a esta concepção: o desejo de autopreservação deve poder funcionar na deliberação e decisão como um desejo de segunda ordem, isto é, uma preferência pela satisfação de alguns desejos em detrimento de outros. Deve-se pressupor que tenha uma força suficiente para ser superior a outros desejos de mesma ordem (as paixões podem operar como *preferências*), que competem pelo papel de ordenação de prioridades de satisfação, ou a desejos imotivados (destes, os desejos e aversões em excesso, como é o caso do “apetite brutal” ou “desejo forte” - manifestado como uma

<sup>31</sup> *An enquiry concerning the principles of morals*, segundo apêndice; p. 172 na trad. argentina.

esse é o caso de todos os desejos de autopreservação

forma de embriaguez ou loucura - característico das "tentações" da avareza, da ambição, da concupiscência e do orgulho, são o alvo principal da psicologia moral desenvolvida por Hobbes),<sup>32</sup> que são formados independentemente dele.

### 1.8. A autopreservação como fim supremo.

Segundo a concepção aristotélica, é com relação a um tal fim ou meta - o bem para o homem ou o fim universal da vida humana - que se definem as duas tarefas básicas da teoria ética: determinar, com a precisão adequada, em que consiste tal fim (ou meta) e qual é a melhor maneira de conseguir realizá-lo (e isto quer dizer duas coisas: qual é a maneira racional de realizá-lo e em que consiste fazê-lo virtuosamente). Estas tarefas são nucleares na construção da teoria ética de Hobbes.

A hipótese sobre a natureza humana exposta em 1.7 também pode significar: existe um fim ou meta pelo qual todas as ações particulares - com seus fins particulares - são feitas, ao qual todas elas visam. E há, para ela, duas interpretações possíveis. Podemos interpretá-la como uma sentença descritiva, através da qual afirma-se que os homens de fato buscam tal fim, ou que é por alcançá-lo que os homens ficam realmente satisfeitos, ou ainda, que este fim permite, quando aceito como hipótese explicativa, de algum modo unificar uma gama extraordinariamente variada e mesmo conflitante de esforços, comportamentos e propósitos detectáveis nas diferentes atividades humanas.

Por outro lado, quando interpretada normativamente, tal hipótese serve como uma prescrição do que deve ser considerado como o fim *próprio* do homem, aquilo que ele *deve* esforçar-se por realizar, o fim ou valor que *deve* motivar ou guiar todas as suas ações.<sup>33</sup> Veremos adiante se é possível localizar estes dois níveis de interpretação na teoria ética de Hobbes e, se este for o caso, a que hipóteses teóricas nela correspondem.

Por enquanto, é importante dizer que, sendo a interpretação hobbesiana do conceito de "fim supremo" ("supremo" é um predicado com significado normativo), a

<sup>32</sup> Cf. *Leviathan*, 8, 36 e 14, 70; e *De cive*, III, XXVI, 116 e XXXII, 120. Este último número indica a página na edição *Sirey*. As citações desta obra, quando não houver outra indicação, são traduções minhas.

<sup>33</sup> Cf. Mackie, *Ethics: inventing the right and the wrong*, pp. 46-48.

autopreservação - entendida como a continuidade do prazer - tem uma função lógica peculiar na *justificação das ações*: funciona como razão última, propósito último ao qual é possível apelar para saber o quanto alguém tem razão em fazer o que faz.<sup>34</sup> No caso de Hobbes, as ações que se provam serem irracionais o são, em última análise, pela razão de não poderem ser meios para a realização do "fim supremo".<sup>35</sup> O sentido deste termo não pode incluir a concepção de que existe algum desejo (ou *função*) cuja plena satisfação (ou *perfeição*) representa a *felicidade* ou *bem supremo*, entendida como um estado no qual um acréscimo ou perda de satisfação futura não estão incluídos (através de um cômputo) em sua definição (ver também 3.3).

7  
 Os  
 Nas  
 justificada?

### 1.9. A autopreservação como fim do "auto-interesse de longo prazo".

Segundo a concepção hobbesiana da ética, o motivo fundamental para agir moralmente (isto é, para alguém se motivar a agir obedecendo aos princípios éticos) é o desejo de *autopreservação* - com esta, veremos o porquê, sendo interpretada corretamente como meta de um "auto-interesse de longo prazo" (*long-term self-interest*).<sup>36</sup> No sistema ético de Hobbes, não há um auto-interesse diretamente racional, assim como não há outros desejos racionais. O que pode haver é uma "busca deliberada e racional" do auto-interesse. O que a filosofia moral hobbesiana pode demonstrar é que, na suposição de que o auto-interesse, corretamente compreendido, é sempre de longo-prazo (e isto é também uma parte do "conhecimento do que é a natureza humana"),<sup>37</sup> no caso de a maior parte das pessoas ("seres humanos") agir moralmente (isto é, se cada uma delas seguir os preceitos morais expressos nas "leis da natureza"), a vida que resulta desta adesão coletiva é a mais racional, isto é, a que melhor serve ao interesse próprio de cada uma delas. Sob tais condições, ser

<sup>34</sup> Cf. *De homine*, 7, 5-6.

<sup>35</sup> Aqui cabe uma observação sobre o significado dos termos "bem" e "fim". Hobbes afirma: "*bonum* e *finis* são nomes diferentes para diferentes considerações da mesma coisa". *De homine*, 7, 5. Apresento a seguinte interpretação desta passagem. Um *bem* diz-se do que é desejado (porque cremos que nos dará prazer); um *fim* diz-se da satisfação do desejo, da "fruição ou uso" do que é desejado. A "mesma coisa", que está sendo tomada sob "diferentes considerações", é precisamente o que descrevemos como "o que é desejado".

<sup>36</sup> Ver, por exemplo, James Sterba, *Justifying morality: the right and the wrong ways*, p.46.

<sup>37</sup> *De homine*, 1, 1.

uma pessoa moral implica em poder esperar a maior felicidade, quer dizer, garantir uma estrutura cooperativa mínima para que uma pessoa possa razoavelmente esperar a continuidade da satisfação dos seus desejos.

Tal desejo é apenas pressuposto pela ética, ou seja, o sistema ético de Hobbes não oferece uma explicação de como é possível que o *desejo de autopreservação* (ou, a sua outra "face", o medo da morte violenta) funcione como um tipo de motivo necessário nos seres humanos. A metáfora da outra "face" pode ser enganadora. O medo da morte violenta só pode predispor "os homens para a obediência ao poder comum" se eles preferirem (i. e., desejarem com mais intensidade) uma condição na qual, graças à segurança e ao conforto, podem melhor cuidar da sua própria conservação e garantir uma vida mais satisfeita a uma outra condição, a saber, a "condição de guerra", na qual podem contar somente com o esforço e trabalho próprios para garantir uma vida futura satisfeita.

Além disso, não é necessário que um indivíduo que tema a morte violenta deseje garantir a sua autopreservação por meio da segurança e do conforto, pois pode ser o caso que um tal indivíduo prefira arriscar-se a obter o que deseja por meio da competição e da luta. Somente o medo da morte violenta "somado" ao "desejo de conforto e deleite sensual ... produz a mesma tendência", a saber, a disposição (paixão) para "escapar" do estado de natureza.<sup>38</sup> Ou, antecipando um aspecto, somente indivíduos que, estando basicamente interessados em sua própria conservação (interessados em continuar existindo), são capazes de raciocinar adequadamente sobre os meios mais eficientes de realizá-la podem também desejar "escapar" do estado de natureza.

<sup>38</sup> Ver *Leviathan*, 11, 47-48 e 13, 63. A "esperança de impunidade" de quem comete uma *injúria* pode dispor sua vontade a desprezar a autoridade do Estado e confiar na *vingança privada* como meio de compensação de um dano sofrido anteriormente. Neste caso, a paixão do *ódio* (não raro associada à *vanglória*, a qual pode levar alguém a sobrestimar a importância do dano que lhe foi causado) é capaz de produzir um crime grave: o assassinato praticado com premeditação. Cf. *Leviathan*, 27, 154-155.

### 1.10. O desejo de autopreservação visto como um desejo categórico.

Gostaria, agora, de introduzir o conceito de *desejo categórico*, tal como foi apresentado por Bernard Williams. Meu principal objetivo é usá-lo como elemento básico da interpretação do papel do desejo de autopreservação na motivação humana. Williams pretende oferecer uma resposta possível ao seguinte problema: por que é preciso que consideremos a morte como um mal? Num sentido mais evidente, pelo fato de que, desejando obter uma coisa qualquer, estamos motivados a opor resistência àquilo que nos impedir de obtê-la; e, a menos que se deseje algo que esteja para além da vida, a morte representa um grande mal. E quando desejamos algo e somos impelidos a tentar obtê-lo, geralmente o fazemos com a suposição de que (embora nem sempre isso venha à consciência) continuaremos vivendo pelo menos o suficiente para vir a desfrutar do que for obtido.

Mas, para desejarmos algo no presente, será sempre necessário condicionar nosso desejo à perspectiva da continuação da nossa própria vida num dado momento futuro? Pensemos no caso de um suicida. Ele pode ponderar racionalmente sobre o seu suicídio, querendo escolher deliberadamente se deve continuar a viver ou não, e decidir não se matar. Concebendo-o numa situação dramática, pensemos que ele, na verdade, decidiu continuar vivendo jogando impulsivamente (ainda que possa ter sido movido por um desejo nebuloso e incipiente) seu revólver num rio. Neste caso, seria absurdo condicionar o seu desejo de continuar vivendo à sua expectativa de continuar vivendo, pois só pode esperar continuar vivendo depois de decidir continuar vivendo. Este é o tipo de desejo que Williams denomina de *categórico*. Ele é categórico no sentido justamente de servir para evitar que a existência de alguém seja colocada em questão ou, para responder ao dilema enfrentado por quem vem a colocá-la em questão. Quer dizer, no sentido de não estar condicionado à existência futura de alguém, mas sim de constituir-se em condição (um motivo, razão ou interesse) para que este alguém exista no futuro.<sup>39</sup>

O desejo de autopreservação pode ser visto como um desejo categórico; em Hobbes, esta consideração irá ajudar-nos a compreender o papel *dominante* daquele desejo. Um problema importante será o de saber se este papel pode ser adequadamente explicado

<sup>39</sup> Ver o artigo *Persons, character and morality in Moral luck*, pp. 10-12.

como resultado da operação de algum mecanismo ou se, tal como é característico em algum momento da investigação científica, não se pode explicar por que aquele desejo é dominante. Quer dizer, se o "homem é uma parte da natureza", por que não tentar compreender a ação humana, e mais particularmente o seu domínio moral, por meio de uma explicação naturalista que seja satisfatória ainda que apenas supondo a *dominância* do desejo de autopreservação, sem oferecer uma prova apriorística da necessidade dele?

Assim, por que não considerar o desejo de autopreservação como uma hipótese valiosa na descrição do comportamento humano (pelo menos aquele referente à esfera moral da vida humana)? Podemos por ora aceitar uma observação de Williams (embora deslocando-a um pouco do seu contexto de argumentação):

*"Meus projetos presentes são a condição da minha existência, no sentido de que a menos que eu seja impulsionado adiante pelo conatus do desejo, projeto e interesse, não fica claro por que deveria, afinal de contas, continuar vivendo ... Os desejos categóricos que nos impulsionam adiante nem mesmo têm de estar manifestos para a consciência, muito menos serem grandiosos ou de grande alcance; um bom testemunho de que a existência de alguém é algo urgente é o fato de sua urgência não ser colocada em questão."*<sup>40</sup>

### [1.11. Tipos e níveis de prova na teoria ética hobbesiana.] alt.

A presença deste componente nuclear da motivação humana pode ser compreendida por meio de uma explicação independente da ética; e isto é o que faz a chamada *teoria da motivação*, que em Hobbes não é uma teoria empírica, mas uma teoria considerada, por boa parte dos comentadores, como inteiramente derivada de uma metafísica geral por meio de raciocínios dedutivos. Porém, o problema está em saber se tal derivação é compatível com o tipo de interpretação tomada como a mais adequada do desejo de autopreservação.

Seja por meio de uma derivação, seja partindo de uma pressuposição forte, aos resultados da sua teoria motivacional aplica-se a propriedade da *universalidade*. Mais especificamente, sem se pressupor a universalidade do desejo de autopreservação (que é um desejo não-racional), não se pode provar, dentro da teoria ética de Hobbes, que os

---

<sup>40</sup> *Op. cit.*, p. 12.

princípios éticos indicam ações cuja realização é uma *necessidade racional*. No próximo capítulo examinarei mais detalhadamente alguns problemas envolvidos na construção da teoria motivacional hobbesiana.

Internamente a sua teoria ética, Hobbes pode provar por que, para indivíduos racionais e informados adequadamente das condições em que se encontram junto com outros indivíduos, agir seguindo as prescrições indicadas pelos princípios (isto é as chamadas "leis da natureza") é *melhor* (isto é, é *racional*) para cada um deles. Este tipo de prova é o que denominamos de *justificação dos princípios éticos* (tratada no cap. 5).

Mas se o que queremos é uma *validação* dos princípios, tal prova não é suficiente. Pois estes só podem ter força - e para isso precisam ser adequados para ingressar na cadeia de motivos que originam as ações - sobre aqueles indivíduos que os aceitam como diretivas que indicam os meios mais eficazes na obtenção do que mais desejam, a saber, a autopreservação.

Para prosseguir, necessitamos introduzir uma distinção que é essencial para a elucidação de muitas questões em ética: a distinção kantiana entre *imperativos categóricos* e *hipotéticos*. Para isso, podemos usar o conceito de "razão para aceitar um imperativo como prescrição" e o de "motivo". Sempre que a razão para um agente aceitar a ação prescrita ou exigida por um imperativo residir em algum motivo que o agente supostamente tem, isto é, residir no fato de a ação prescrita ser um meio disponível para realizar pelo menos um dos fins por ele desejado, então tal imperativo é hipotético e a razão, prudencial. Não satisfeita esta condição, o imperativo é categórico.

Os imperativos morais hobbesianos têm a sua validade condicionada à prévia existência de um desejo; por isso são denominados de *imperativos hipotéticos*. Neste caso, em oposição à concepção kantiana, os princípios morais (quer dizer, regras gerais que justificam a conduta moral) da teoria de Hobbes não podem ser considerados como fonte absoluta da motivação para agir moralmente.<sup>41</sup> Isto é, seguindo Hobbes, estes princípios não são válidos independentemente de qualquer hipótese, dada previamente, sobre a motivação humana. Por sua vez, o que Kant não aceita é precisamente que, dentre os motivos para agir moralmente, possa haver algum não puramente racional, isto é, algum motivo não inteiramente derivado dos próprios princípios éticos.

---

<sup>41</sup> Gostaria de esclarecer que, quando falo de "princípio", não estou fazendo uso da distinção kantiana entre "lei ou princípio" e "máxima".



### 1.12. Acordo moral e contrato social.

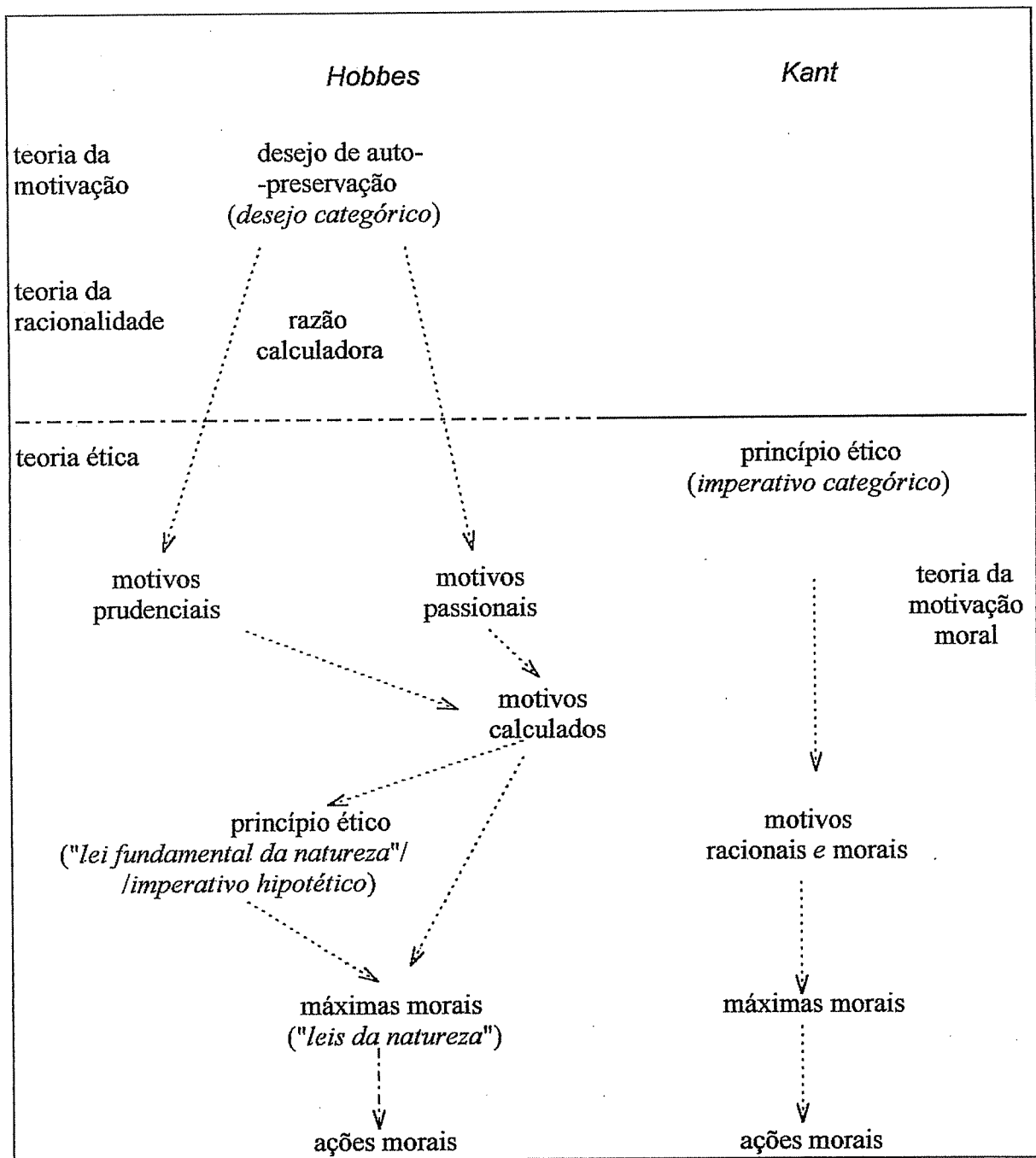
Talvez o problema nuclear a ser solucionado por qualquer teoria do contrato social seja o do acordo para o estabelecimento de regras morais necessárias para a realização da *cooperação social* (que não podem ser somente regras de não-agressão). Esta pode ser representada pelo chamados *termos da cooperação*, que nada mais são que as regras acordadas entre indivíduos livres e iguais, que estabelecem genericamente os limites do que é justo ou não.

O que pode ser teoricamente aguardado como resultado da adesão coletiva das partes contratantes - bem como a prova da estabilidade de um tal acordo - depende da caracterização da *situação pré-contratual*, isto é, da descrição da situação hipotética na qual se situam indivíduos com uma dada psicologia (disposições, preferências, grau de racionalidade, etc.) que buscam realizar seus desejos sem nenhuma restrição comportamental além das limitações oferecidas pela oposição direta de outros indivíduos. Esta é portanto uma oposição entre *indivíduos livres* - indivíduos que não detêm qualquer "direito de mando [uns] sobre os outros" - que dispõem de *iguais poderes* para garantir-lhes a conservação e defesa de si próprios.<sup>42</sup>

---

<sup>42</sup> Cf. *Leviathan*, 15, 80.

*Esquema comparativo: arquitetônicas das teorias éticas hobbesiana e kantiana.*



## 2. TEORIA DA MOTIVAÇÃO

*"A alma ... esforça-se por perseverar no seu ser por uma duração indefinida, e é consciente do seu esforço"*

Baruch Spinoza, *Ética*

*"Sem estados corporais que a seguem, a percepção sensorial teria uma forma puramente cognoscitiva, pálida, ... sem calor emocional. Poderíamos então ver ursos e julgar a fuga a propósito, receber uma ofensa e julgar ser bom bater, mas não experimentaríamos realmente nem temor nem cólera."*

William James, *The principles of psychology*

### 2.1. O esquema conceitual da "psicologia materialista".

Neste capítulo, apresento uma explicação (no sentido de *construção racional*) do significado do desejo de autopreservação - que está na base da teoria da motivação de Hobbes - construída como uma derivação conceitualmente interna à "psicologia materialista" (termo cujo significado será analisado no decorrer do argumento). Também estou interessado em apontar algumas limitações dessa explicação.

Para tanto, podemos começar condensando as definições de "desejo" e "aversão" que aparecem no *Leviathan*, apresentando-as como se segue:

Desejo (ou Appetite) = def os pequenos começos do movimento - quando este é em direção ao objeto que o causa - dentro do corpo do homem, antes das ações visíveis.

Aversão = def os pequenos começos do movimento - quando este é de distanciamento relativamente ao objeto que o causa - dentro do corpo do homem, antes das ações visíveis.<sup>43</sup>

*Desejo e aversão* são tipos de *movimento vital* (movimento que ocorre "dentro do corpo do homem"), que dá origem a toda e qualquer ação visível. Este movimento é causado pelos objetos (i. e., "corpos exteriores") que afetam os sentidos. Estamos descrevendo uma relação de *causação* entre objetos externos e nossos corpos. Mais ainda,

---

<sup>43</sup> Cf. *Leviathan*, 6, 23.

estamos falando de fenômenos descritos comumente pela psicologia; no entanto, definindo alguns dos seus conceitos básicos usando somente termos pertencentes à linguagem da física: partículas fundamentais constituintes dos nossos corpos são movimentadas, segundo leis determinísticas, por outras partículas, a saber, aquelas constituintes dos objetos externos aos nossos corpos.

Desta forma, podemos fazer um axioma da física ingressar como axioma de uma "teoria psicológica" (entendida como uma teoria que serve inicialmente para descrever eventos mentais) e depois reconstruímos tal psicologia traduzindo a sua linguagem para a linguagem da mecânica. O objetivo básico desta tentativa de tradução é buscar explicar as ações humanas inteiramente como efeitos de movimentos de partículas sujeitas a leis mecânicas, movimentos que supostamente descrevem os motivos que as causam.

Podemos aplicar a este projeto hobbesiano uma observação feita por Stuart Hampshire sobre a fonte da teoria da mente de Spinoza, porque Hobbes também foi, neste aspecto, fortemente influenciado pelo "novo iluminismo científico" que marcou o século XVII:

*"Da mesma forma que a teoria física revela um novo mundo de partículas em movimento por trás do mundo usualmente percebido de objetos de média escala, a psicologia e a filosofia da mente revelam uma nova realidade psíquica por trás das paixões usualmente percebidas e dos nossos propósitos usuais."*<sup>44</sup>

É possível ver Hobbes como um dos "pioneiros em teoria psicológica", expressão que designa um grupo de pensadores cujos esforços teóricos marcaram as primeiras tentativas de explicar a motivação e a conduta humanas como efeitos de regularidades na operação de um mecanismo mental, regularidades inteiramente descritas por leis da dinâmica, quer dizer, por leis do movimento válidas para explicar (e prever) padrões de interação entre corpos. Gilbert Ryle indicou acuradamente alguns pressupostos e implicações deste ambicioso projeto - dominante através de variações que se estenderam até o início do séc. XX - de mapear um conjunto de forças físicas básicas que explicassem toda a conduta humana. Podemos considerar as suas observações como reveladoras de motivações incluídas no projeto hobbesiano, assumindo que o contexto científico de

---

<sup>44</sup> *Morality and conflict*, p. 10.

Hobbes era basicamente o da física galilaica e que - o que é muito importante - seu projeto era *reducionista*:

*"Pensando em sua missão científica como a de copiar para o mundo mental o que os físicos tinham feito para o mundo material, procuravam contrapartidas mentais das forças em cujos termos foram dadas explicações dinâmicas dos movimentos dos corpos. Que fenômenos passíveis de introspecção fariam pela conduta humana intencional o que a pressão, o impacto, a fricção e a atração fazem pelas acelerações e desacelerações de objetos físicos? Desejo e prazer, aversão e dor, pareciam admiravelmente qualificados para desempenhar os papéis requeridos; tanto mais que, como é de conhecimento comum, as pessoas normalmente querem o que terão prazer em ter quando o obtiverem, normalmente ficam satisfeitas por obter o que queriam ter, e, quando escolhem entre uma coisa e uma outra, preferem a coisa que escolhem à que rejeitam. Esses estados de espírito são variáveis tanto em grau quanto em duração. Seria de se esperar que algo como o paralelogramo de forças se aplicasse aos nossos gostos e desgostos, desejos e aversões, convergentes e divergentes."*<sup>45</sup>

## 2.2. Uma linguagem fisicalista para a psicologia.

Os passos de uma possível redefinição sistemática da linguagem da psicologia estão esquematizados na seguinte seqüência de afirmações (extraídas do *Leviathan* e *De homine*), que são axiomas e teoremas para a descrição da *realidade*, i. e., das coisas como elas são, e suas contrapartidas usadas na descrição da *aparência* da realidade, i. e., das coisas tal como nos aparecem (ou, tal como as concebemos):

[*Realidade*]

- i) Axiomas: i.1. "o movimento nada produz senão o movimento" (1, 3);
  - i.2. o que realmente causa a sensação é o movimento das coisas fora de nós (cf. *De homine*, 2, 2 e 10);
- ii) "são muitos os movimentos da matéria que pressionam nossos órgãos de maneira diversa" (1, 3);
- iii) "a causa da sensação é o corpo exterior, ou objeto, que pressiona o órgão próprio de cada sentido" (1, 3);
- iv) "(...) na sensação, aquilo que realmente está dentro de nós é apenas movimento (...) provocado pela ação dos objetos externos" (6, 25);

<sup>45</sup> *Dilemmas*; p. 91 na tradução brasileira.

v) "[as qualidades denominadas sensíveis, i.e, luz (ou *cor figurada*), som cheiro, sabor, frio, calor, dureza, macieza, etc.] nada mais são do que movimentos diversos" (1, 3);

vi) "[a imaginação é apenas o resíduo] do movimento provocado nos órgãos e partes inferiores do corpo do homem pela ação das coisas que vemos, ouvimos, etc." (6, 23);

vii) "a imaginação é a primeira origem interna de todos os movimentos voluntários" (6, 23);

viii) "o efeito (...) realmente produzido [pela ação dos objetos externos] não passa de movimento, ou esforço, que consiste em apetite ou aversão em relação ao objeto" (6, 25).

[*Aparência (ou representação da realidade)*]<sup>46</sup>

i) Axioma: "Há continuamente nas nossas mentes certas imagens ou concepções das coisas fora de nós" (1, 8). "Imagens ou concepções" sendo definidas como "aparições 'em nós' do movimento provocado pelo objeto 'no cérebro ou nos espíritos, ou em alguma substância interna da cabeça' " (cf. 1, 8 e 2, 4);

ii) "através dos nossos órgãos [da sensação] temos diversas concepções de diversas qualidades dos objetos" (2, 3);

iii) "todas as concepções [são produzidas pelas] ações da própria coisa [i. e., o OBJETO da sensação] de onde procede a concepção" (2, 2);

iv) "quando a ação é presente, a concepção que essa ação produz chama-se SENSACÃO" (2, 2);

v) as sensações são concepções, de diversas qualidades ou naturezas dos seus objetos, inerentes não ao objeto, mas àquele que sente (cf. 2, 3 e 4);

vi e vii) "a imaginação [é] uma concepção remanescente" (obscurecida e enfraquecida pela sensação de outros objetos presentes ou, como pode acontecer durante o sono, clara e forte graças à "privação do ato de sensação" (cf. 3, 1 e 2).

A redução da linguagem da psicologia a uma linguagem fisicalista (na qual se pressupõe que apenas partículas dotadas ou não de movimento existem) é o que nos autoriza a falar de uma "psicologia materialista". Segundo Jean Hampton, esta também funciona como "base fundacional" da teoria do valor de Hobbes.<sup>47</sup> Para Hobbes existe

<sup>46</sup> Todas as citações a seguir são do *De homine*.

<sup>47</sup> Ver *Hobbes and the social contract tradition*, pp. 11-19.

somente um único mundo real. Ainda que se possa descrevê-lo usando diferentes sistemas simbólicos ou linguagens, a ontologia da linguagem da física não é relativa a esta linguagem, mas a única possível.

### 2.3. O "materialismo metafísico".

Conseqüentemente, o projeto da "psicologia materialista" faz parte de uma defesa ampla do chamado *materialismo metafísico*, do qual Hobbes é por vezes considerado o fundador moderno. De acordo com esta doutrina geral, a linguagem da física é suficiente para a construção de uma descrição completa de tudo o que há. Um dos seus enunciados possíveis está expresso no seguinte trecho:

*"O mundo (não quero dizer apenas a terra, que denomina aqueles que a amam homens mundanos, mas também o universo, isto é, toda a massa de todas as coisas que são) é corpóreo, isto é, corpo, e tem as dimensões de grandeza, a saber, comprimento, largura e profundidade; também qualquer parte do corpo é igualmente corpo e tem as mesmas dimensões, e conseqüentemente qualquer parte do universo é corpo e aquilo que não é corpo não é parte do universo. E porque o universo é tudo, aquilo que não é parte dele não é nada, e conseqüentemente está em nenhures"<sup>48</sup>*

A idéia de um universo homogêneo, cujas partes têm todas *a mesma natureza*, constitui-se numa ruptura com a concepção aristotélica. Temos assim o estabelecimento de uma nova concepção geral da realidade física do universo - uma explicação metafísica (*philosophia prima*) sobre a natureza dos corpos, do espaço, da matéria, do movimento, em resumo, "o estabelecimento do significado" dos princípios ou causas primeiras dos eventos naturais. O mundo aristotélico é um mundo composto por um grande número de indivíduos com "naturezas" bem determinadas, específicas, que formam um todo bem ordenado, um *cosmos*, no qual cada um ocupa um lugar próprio, conforme a sua natureza (seu "lugar natural").

No mundo hobbesiano, não existem diferentes propriedades correspondendo (figurativamente ou não) às diferentes sensações (visuais, tácteis, olfativas, etc.) de um

---

<sup>48</sup> *Leviathan*, 46, 371.

sujeito. O mundo é realmente muito diferente de como o percebemos: pode-se dizer que a essência das coisas materiais ou corpóreas (que são as partes de um "corpo indefinidamente extenso", o universo, nos termos de Descartes)<sup>49</sup> é a sua tridimensionalidade ("extensão" é o termo cartesiano). São os movimentos dos corpos (ou de suas partes) que causam as nossas diferentes sensações e, conseqüentemente, suas idéias correspondentes.

Para Hobbes, no entanto, não pode haver espaço sem corpo, simplesmente porque não pode existir algo como uma *essência separada*: "... a essência, enquanto distinta da existência, não é senão uma reunião de nomes pelo verbo *é*; e portanto, a essência sem a existência é uma ficção do nosso espírito."<sup>50</sup> Adaptando um exemplo apresentado por ele, poderíamos portanto sustentar que quando dizemos que "o corpo é uma coisa *extensa* (ou *espacial*)" estamos identificando *extensão* e *corpo*; contudo, não queremos com isso afirmar que "na realidade, a mesma extensão ... que constitui o espaço, constitui o corpo".<sup>51</sup> Isto porque, ainda que os termos *coisa extensa* (*res extensa*) e *corpo* (*corpus*) signifiquem a mesma coisa, a saber, a substância corpórea, o mesmo não ocorre quando afirmamos que "um corpo é espacial" ("uma coisa extensa é espacial"), pois neste caso o que queremos afirmar é tão somente que o nome *espacial* é necessariamente conseqüente ao nome *corpo*.

O termo *espacial* não pode significar algo real (*entitas*), o que seria denominado como uma "essência real" ou, como denominavam os escolásticos, uma "forma substancial" - isto é, algo existente tanto na coisa quanto no pensamento, porque os termos *entidade*, *essência*, *essencial*, *essencialidade* ... [assim como muitos outros que derivam do verbo *esse* e suas conjugações] não são "nomes de coisas, mas sinais pelos quais tornamos conhecido que concebemos a conseqüência de um nome ou atributo em relação a outro"<sup>52</sup>

O nome *corpo*, segundo Hobbes, nomeia tudo o que é real, quer dizer, tudo o que é uma coisa (*res*) ou substância (*substantia*). Segue-se que, para Hobbes, só há um tipo de substância: a *substância corpórea*. Segue-se também que tudo que é *real* é *corpóreo*. E

<sup>49</sup> *Principia philosophiae*, parte II, parágrafo 21. Citado por Alexandre Koyré, *Do mundo fechado ao universo infinito*, p. 105. Nesta obra, o autor faz uma apresentação algo detalhada dos conceitos de mundo, matéria e espaço (e suas relações) em Descartes. Ver principalmente pp. 101-110.

<sup>50</sup> *Terceiras objeções às Meditações de Descartes, Objeção décimo-quarta*.

<sup>51</sup> Identificação que Descartes fez, por exemplo, nos *Principia philosophiae*, parte II, 10. Citado por Alexandre Koyré, *op. cit.*, p. 104.

<sup>52</sup> *Leviathan*, 46, 372.



assim como não faz sentido dizer que algo é mais corpóreo que outro algo (pode alguma coisa ser dita mais ou menos tridimensional?), igualmente não faz sentido dizer que algo tem mais realidade (ou, melhor dito, mais *graus de realidade*) que outro algo.<sup>53</sup> Segue-se

---

<sup>53</sup> Na teorização filosófica hobbesiana sobre a linguagem (ou, como atualmente sói ser denominada, a "filosofia da linguagem" de Hobbes), toda palavra é um sinal que serve para nomear algo. Para uma palavra ter significado, é preciso que seja (ou possa ser definida como) *nome* de algo. Se uma palavra não é um nome, é porque sua função na sentença não é a de denotar algum *objeto* (ver abaixo o sentido deste termo), mas indicar alguma operação (lógica) que fazemos para raciocinar, isto é, para fazer um cálculo das conseqüências de nomes. Ver *Leviathan*, 4, 14 e 46, 372.

Ao introduzir o conceito de nome como única categoria semântica - i. e., como conceito que descreve o único tipo de vínculo ou relação representativa possível entre nossas palavras e os objetos na realidade - Hobbes não está fazendo uma reflexão original sobre a natureza da linguagem. Está, ao contrário, sendo tributário de um tipo de investigação filosófica hoje denominada "concepção agostiniana da linguagem". Santo Agostinho defende a tese de que ainda que as palavras cumpram funções sintáticas diferentes numa oração, se elas são significativas, então são nomes, funcionam como nomes. No diálogo *De magistro*, quando Agostinho pede a seu interlocutor, Adeodato, que explique o significado de duas determinadas palavras, ele lhe solicita definições ostensivas: "gostaria que, se pudesses, me mostrasse as coisas mesmas de que estas [palavras] são sinais" *De magistro*, II, 4.

O objeto (uma *coisa*, i. e., um indivíduo ou *particular*; um *universal*, i. e., um *nome comum* a muitas *coisas* (ou *particulares*) - semelhantes em pelo menos alguma qualidade essencial ou accidental - ou ainda uma *imagem mental*, i. e., pensamentos, concepções e imaginações - dentre as quais Hobbes inclui as paixões) representado pela palavra é o seu significado. Duas palavras têm o mesmo significado se representam o mesmo objeto. Compreendemos uma palavra (ou um *discurso*, i. e., uma *cadeia* de palavras) se associamos mentalmente a palavra com o objeto de que é o nome. Se não logramos fazer uma tal associação, não compreendemos as palavras que vemos ou ouvimos; elas são, por isso, meras marcas no papel ou sons insignificantes. Esta concepção de *compreensão* como um processo mental também é encontrada em Santo Agostinho: "no que diz respeito a todas as coisas que compreendemos, não consultamos a voz de quem fala, a qual soa por fora, mas a verdade que dentro de nós preside à própria mente, incitados talvez pelas palavras a consultá-la" *Op. cit.*, XI, 38.

Um *nome próprio* (ou *singular*) significa somente um indivíduo ("traz ao espírito uma coisa [particular] apenas"); um nome universal é atribuído a muitas "coisas particulares" (*indivíduos*), trazendo ao espírito - representando, significando, nomeando, referindo -, a cada vez que é compreendido, qualquer uma, e apenas uma por vez, dessas muitas "coisas particulares" (*indivíduos*). Quanto maior o número de coisas às quais é atribuído um nome universal, maior a sua "extensão" e, *ipso facto*, maior ou mais ampla a sua "significação" Cf. *Leviathan*, 3 e 4.

Como pode ser observado, não há a importante distinção lógica estabelecida por Frege entre *sentido* e *referência* de um termo; indistinção que tende a gerar conseqüências paradoxais. Em que pese esta falha teórica, Hobbes é considerado como um dos "fundadores da filosofia da linguagem", principalmente por sua resposta radical àquele que talvez seja o problema básico deste campo de investigação filosófica: como a linguagem contribui para o funcionamento de nossas faculdades cognoscitivas e, por conseqüência, para a nossa conceitualização da realidade? Para uma apresentação sucinta da resposta hobbesiana a este problema, ver Eduardo Rabossi, *Hobbes y la filosofía del lenguaje*.

disso que o nome *substância incorpórea* é uma contradição em termos, um nome que "não significa absolutamente nada".<sup>54</sup>

Uma visão sistemática da natureza deve incluir a explicação das mudanças por que passam os corpos. Um mesmo corpo só muda se algum dos seus acidentes muda. Uma mudança física importante é aquela resultante do próprio movimento dos corpos. Na física aristotélica, as coisas têm o seu "lugar natural" e estariam "em ordem" caso se encontrassem cada uma em seu lugar, no qual permaneceriam em repouso. A ordem é concebida como estática, um estado de equilíbrio cuja ruptura é que demanda explicação. Alexandre Koyré observa que, para esta teoria, "não há necessidade de explicar o repouso, pelo menos o repouso natural de um corpo em seu lugar próprio; é a sua própria natureza que o explica, que explica por exemplo, o repouso da Terra no centro do mundo".<sup>55</sup>

A física de Hobbes tem dois pilares metafísicos: uma teoria mecanicista e outra, determinista. Um acontecimento - uma mudança - no interior do mundo físico pode ser totalmente descrito em termos de matéria extensa em movimento, cuja causa é sempre um *empurrão*. Ao mesmo tempo, cada acontecimento é totalmente determinado ou causado por acontecimentos passados (que são as condições iniciais) operando em cadeia por meio de leis determinísticas.

A chamada "psicologia materialista" é uma base possível da teoria que gera as explicações sobre as motivações ou inclinações humanas, e assim deve servir também para

<sup>54</sup> Ver *Leviathan*, 4, 17 e 46, 371-374. A chamada "tabela dos absurdos", ou seja, um sumário que Hobbes faz de todas as combinações de nomes que resultam em proposições absurdas, supõe, não por acaso, uma classificação dos nomes correspondente às "cinco unidades básicas da sua filosofia". Os absurdos são gerados sempre que, numa proposição, são copulados nomes de objetos ontologicamente diferentes. Para S. Morris Engel, a tabela é construída a partir de considerações não-lingüísticas e é visada para um propósito extra-lingüístico. Hobbes, defende este autor, construiu tal tabela não a partir de "*insights independentes*", mas com o propósito de consolidar resultados obtidos em dois domínios basilares do seu sistema filosófico: sua *teoria dos universais* (uma ontologia) e sua *teoria da percepção* (uma epistemologia). Engel conclui seu artigo com extrema clareza:

"A 'tabela dos absurdos' construída por Hobbes é uma tentativa de emprestar um peso adicional a um corpo de pensamento ao subsumi-lo sob um plano esquemático geral com a sanção e a autoridade da linguagem. Ela não mostra ou prova nada sobre a lógica da linguagem que aquele corpo de pensamento não tenha suposto já haver provado e, portanto, não descobre nada sobre a linguagem que já não tenha sido pressuposto por aquela filosofia. Se ela é uma 'técnica de análise lógica', então tal técnica é adequada somente para a filosofia de Hobbes, ou para uma filosofia que aceita as suas pressuposições". *Hobbes's "table of absurdity"*, p. 542.

<sup>55</sup> *Estudos galilaicos*, pp. 23-24.

determinar o significado do desejo de autopreservação. Deste modo, também é possível estabelecer uma *teoria da motivação* por meio de raciocínios *a priori*.<sup>56</sup>

#### 2.4. Construindo uma teoria da motivação com conceitos da "psicologia materialista".

Façamos agora uma incursão pela teoria da motivação que pode ser construída com elementos daquela psicologia. A principal limitação de uma tal teoria decorre, como veremos, precisamente desta construção, ou seja, da tentativa de explicar as motivações humanas pressupostas pela ética usando somente termos pertencentes à *psicologia materialista*.

Um dos principais objetivos deste capítulo é mostrar que um primeiro (e mais fundamental) passo desta redução - a adoção de uma interpretação por assim dizer *materialista* do desejo de autopreservação - resultaria num beco sem saída para Hobbes, isto é, numa impossibilidade de derivar todas as conseqüências que ele pretende. Veremos como e por que a psicologia materialista poderia servir de fundamento do comportamento animal - se para a explicação deste pressupusermos que os animais sempre respondem de um modo automático e míope ao seu ambiente - mas não do comportamento humano. E também como tais diferenças aparecem elas próprias na teoria de Hobbes.

Dentro do esquema conceitual da psicologia materialista, as alterações da vontade são apresentadas como casos de movimento natural, explicáveis por leis da física. Assim, tudo o que parece bom a cada um é objeto da sua vontade, objeto desejado. "Aparência" deve ser entendida como *sensação*; por isso, "parece bom" significa "é sentido como bom", isto é, "o que é sentido como prazeroso". Dito de outra forma, por definição cada um não pode deixar de querer o que lhe parece bom. E isto é uma inclinação tão natural quanto o movimento de queda de uma pedra depois de jogada para o alto.

---

<sup>56</sup> Uma apresentação de uma derivação deste tipo, partindo do chamado "axioma da finitude", interpretado por meio do princípio da inércia de Galileu, pode ser encontrada no início do artigo *Condições de autoridade e autorização em Hobbes*, do professor Balthazar Barbosa.

Igualmente, também é uma inclinação natural de cada um evitar aquilo que lhe parece um mal.<sup>57</sup>

O caráter natural das inclinações da vontade é, em última análise, o que explicaria (no sentido de se poder derivar) a existência de um desejo dominante: o desejo de autopreservação. Esta explicação não pode desconsiderar a concepção hobbesiana de *vontade*. Para Hobbes, a vontade é apenas "o último apetite na deliberação" e as ações não dependem de algo como uma vontade que pode se auto-determinar - pelo simples fato de não existir algo isento da necessidade natural -, mas de uma inclinação que pode ser considerada como uma resultante de desejos que competem com diferentes forças. Isto quer dizer que "a vontade também é necessariamente causada por outras coisas das quais não dispõe; [seguindo-se disso que] as ações voluntárias têm, todas elas, causas necessárias".<sup>58</sup> Portanto, a vontade humana é *livre* somente no sentido de que os homens possuem a faculdade de fazer o que desejam fazer.<sup>59</sup>

<sup>57</sup> Cf. *De cive*, I, I, VII e XIV.

<sup>58</sup> *Quaestiones de libertate et necessitate, contra Doctorem Bramhallum*; citado por Karl Popper, *The open universe*; na trad. portuguesa, p. 38. Hobbes, em sua autobiografia escrita em dísticos latinos, conta ter escrito, em inglês, dois opúsculos contra John Bramhall, bispo de Londonderry. O primeiro é a obra redatada em 1646 e conhecida pelo título incompleto "*On liberty and necessity*". Publicada somente em 1654, foi rebatida por Bramhall no ano seguinte. Em 1656, Hobbes escreve seu segundo opúsculo contra o bispo. Hobbes resume assim esta polêmica: "a única questão debatida era: por arbítrio de quem queremos, pelo nosso ou pelo de Deus? Ele segue a escola, enquanto que meu guia é a razão". Tradução do latim de Miguel Ángel Rodilla; *Diálogo entre un filósofo y un jurista y escritos autobiográficos*, p. 160.

<sup>59</sup> Cf. *Leviathan*, 6, 28 e *De cive*, II, XIV. Esta definição de liberdade é, segundo Kant, incompatível com o conceito de "vontade de um ser racional", somente na qual pode ser encontrado "o bem supremo e incondicionado", único fundamento possível do *dever moral*. A definição (*negativa*) de liberdade que Kant defende como condição necessária da explicação do conceito de "autonomia da vontade" é a seguinte:

"A *vontade* é uma espécie de causalidade dos seres vivos, enquanto racionais, e *liberdade* seria a propriedade desta causalidade, pela qual ela pode ser eficiente, independentemente de causas estranhas que a *determinem*; assim como *necessidade natural* é a propriedade da causalidade de todos os seres irracionais de serem determinados à atividade pela influência de causas estranhas". *Fundamentação da metafísica dos costumes, terceira seção*, BA 97.

Henry Allison faz um exame do que chama a "tese da reciprocidade", isto é, a afirmação de Kant de que "liberdade da vontade e lei moral são conceitos recíprocos" no excelente artigo *Morality and freedom: Kant's reciprocity thesis*. Esta tese é estabelecida por meio de uma *análise metafísica*, isto é, uma análise dos conceitos e de sua relação. Tal análise é apenas a primeira parte da "justificação da moralidade". A segunda etapa constitui-se num argumento *transcendental* que visa provar ao cético moral não que "as formulações do imperativo categórico realmente captam os princípios que se acham por trás dos juízos morais do homem racional comum" (Ralph Walker, *Kant tem uma filosofia moral?*, p. 9), mas que há algo no mundo correspondendo aos conceitos de obrigação, dever, moralidade e liberdade - que estes sejam aplicáveis ou, no jargão kantiano, que tenham "validade objetiva" - e que torna os nossos juízos morais (crenças sobre o que *devemos* - ou estamos

Para dizer que alguém deseja algo, seria suficiente dizer que tal coisa, uma vez obtida, aumenta, por meio de um mecanismo de *feedback* positivo, o movimento vital do seu corpo; por outro lado, se algo causa aversão a alguém, basta descrever este fenômeno psicológico dizendo que tal coisa - por meio de um *feedback* negativo - diminuiu o movimento vital. Este mecanismo de *feedback* é apresentado por Hobbes na seguinte passagem, como indicam as partes que destaco:

*"Este movimento a que se chama appetite ... parece constituir uma corroboração do movimento vital, e uma ajuda prestada a este. Portanto, as coisas que provocam deleite eram, com toda a propriedade, chamadas jucunda (à juvando) porque ajudavam e fortaleciam; e eram chamadas molesta, ofensivas, as que impediam e perturbavam o movimento vital."*<sup>60</sup>

### 2.5. Construção teórica do desejo de autopreservação.

Segundo a psicologia materialista, o *desejo de autopreservação* é fundamental por uma necessidade resultante do próprio funcionamento (normal) da maquinaria do corpo humano. Esta frase telegráfica é uma paráfrase de uma afirmação algo enigmática que Jean Hampton, intérprete contemporânea de Hobbes, usa para comentá-lo:

*"(...) o desejo de autopreservação [ou, como a autora descreve-o, "nosso desejo fundamental de aumentar nossos movimentos corporais"] é ... explicado como uma função da nossa natureza fundamental para buscar prazer e evitar dor."*<sup>61</sup>

Aceitando então esta afirmação como uma pista a ser inicialmente seguida, é preciso tentar oferecer uma explicação mais articulada do modo como esse desejo é o produto final do processo que estaria ocorrendo no corpo descrito como máquina. O corpo humano está construído para funcionar como uma máquina produtora de prazer. Este funcionamento forma um sistema fechado (ilustração abaixo), isto é, um sistema no qual os

---

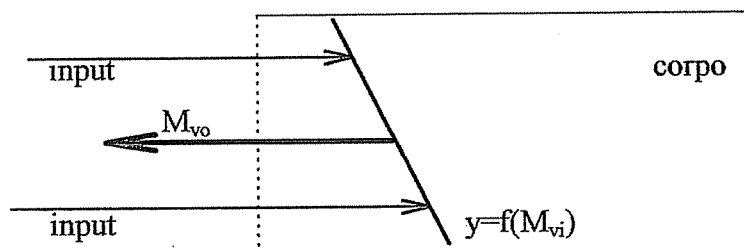
moralmente obrigados a - fazer) objetivos (possíveis de serem verdadeiros ou falsos). Com tal argumento, Kant pretende mostrar ao cético que a moralidade não é uma "idéia quimérica" ou um mero "fantasma do cérebro". Ver Walker, *op. cit.*, principalmente pp. 7-9.

<sup>60</sup> *Leviathan*, 6, 25.

<sup>61</sup> *Hobbes and the social contract tradition*, p. 17. Grifo meu.

*outputs* fazem parte de um mecanismo de retroalimentação (*feedback*), vindo a interferir novamente no mesmo funcionamento e a modificá-lo.

Esquema gráfico do mecanismo de autopreservação



$y$ : função "quantidade de prazer"

$y=f(M_{vi})$

$M_{vo}$ : variável "movimento voluntário"

$M_{vi}$ : variável "movimento vital"

$M_{vi}$ : função "movimento vital"

$M_{vi}=g(i, t)$

$i$ : variável "input"

$t$ : variável "tempo"

Para produzir prazer, o mecanismo básico é o seguinte: a máquina recebe como *inputs* as impressões causadas pelos objetos externos. Tais impressões acionam o mecanismo interno do corpo, que produz, como *outputs*, movimentos voluntários (ações). Se um objeto externo é um objeto capaz de produzir, quando for obtido, o aumento de prazer, formam-se apetites, isto é, movimentos iniciais que se transmitem numa cadeia que termina com os movimentos observáveis do corpo que traduzem o esforço do organismo para obter o objeto. Caso contrário, quando um objeto poderá causar a diminuição do prazer, forma-se uma aversão, ou seja, um tipo de "movimento vital" que se transmite até o final da cadeia, que se torna visível com algum movimento corporal traduzindo o esforço do organismo para evitar tal objeto.

Uma pergunta direta: como o mecanismo "sabe" que "está diante de" um objeto tal que deve produzir movimentos "de apetite" ou "de aversão"? A pista que quero seguir está assim delineada: como posso explicar o desejo de autopreservação como uma função do mecanismo de formação de desejos, que são - considerando a nossa "natureza fundamental para buscar prazer e evitar dor" - efeitos gerados mecanicamente graças

supostamente a algum tipo de informação sobre o aumento de prazer armazenada pelo próprio mecanismo?

A autopreservação não poderia ser descrita como apenas um dentre tantos objetos específicos que são desejados justamente porque causam um aumento do prazer, pois se isto fosse assim, de onde viria então a sua suposta predominância sobre os outros objetos desejados ou, para ser mais preciso, que papel explicativo teria o desejo autopreservação como fundamental na manutenção do mecanismo formador dos movimentos vitais?

Certamente esta predominância não pode ser corroborada se a interpretamos como significando uma operação sempre presente do desejo consciente de prolongar o tempo de vida como ponto de partida de cada movimento particular do corpo. A autopreservação não é predominante no sentido de ser o objeto sempre presente de algum dentre os desejos que explicam as ações particulares. O seu papel na explicação materialista da continuidade do mecanismo de "buscar prazer e evitar dor" é outro.

O desejo de autopreservação deve ser fundamental para a explicação da formação de qualquer desejo particular (incluindo as *aversões*, que podem ser descritas, segundo o próprio Hobbes, como *esforços para evitar algo*). Para tanto, se queremos continuar montando uma explicação de caráter mecanicista, deve ser suficiente partir da concepção de um desejo (de fazer ou de omitir-se de fazer *x*) como um fenômeno inteiramente descrito como um tipo de movimento.<sup>62</sup>

---

<sup>62</sup>Um desejo, sob outra descrição, que não é relevante para a concepção aqui em foco, é uma *intenção*, isto é, um fenômeno mental pelo qual se concebe um fim (ou meta) a ser alcançado, e que só pode ser individualizado com referência a tal fim. Usando desta outra definição, estaríamos produzindo uma explicação de tipo *intencional*.

## 2.6. Um modelo biológico de explicação funcionalista.

Gostaria agora de avançar mais um pouco na pista que sigo. A autopreservação é a *função* do mecanismo produtor de prazer, ou ainda, *o fim ou propósito a que se presta um tal mecanismo*. A estrutura da explicação em questão é a seguinte: o mecanismo gera, como seu efeito, a autopreservação, a qual, por meio de um circuito de retroalimentação causal, mantém aquele mecanismo. O importante é que a retroalimentação opera por meio de um efeito do mecanismo cuja própria persistência no tempo é o objeto da explicação. Esquemáticamente: (i) a autopreservação é um efeito do funcionamento do mecanismo produtor de prazer; (ii) a autopreservação é o que explica causalmente a própria continuidade do mecanismo de formação dos desejos particulares. Este tipo de esquema, próprio da explicação funcionalista, parece ser o que Hobbes tem em mente num trecho aparentemente enigmático:

*"[que uma coisa seja desejada] é um sinal da sua verdadeira necessidade, ou uma prova verossímil de sua utilidade para a conservação daquele que a deseja."*<sup>63</sup>

Na tentativa de construir uma hipótese funcionalista sobre o papel do desejo de autopreservação, pode-se partir de uma apresentação esquemática do processo de seleção natural das espécies. De um modo bem simplificado, consideremos o mecanismo de seleção para organismos colocados num ambiente paramétrico, o que significa um ambiente caracterizado por um conjunto fixo de estímulos que podem afetar cada organismo. Afetam-nos produzindo neles, ao acaso, pequenas mutações. Como hipótese simplificadora, pode-se assumir que não podem surgir novidades genéticas através de outras fontes.

Cada organismo pode então ser concebido como uma máquina que dispõe, num momento qualquer do tempo, de critérios bem definidos para aceitar ou rejeitar as

<sup>63</sup> *De cive*, I, X, 85. Este seria exatamente o esquema de uma explicação *funcionalista*, comum originariamente ao terreno da biologia darwiniana. Segundo Stephen Jay Gould, a "essência do darwinismo", isto é, a hipótese da "criatividade da seleção natural" (que tem uma estrutura explicativa funcionalista), pode ser explicitada do seguinte modo:

"No sistema de Darwin, a seleção natural tem um papel construtivo: constrói aos poucos a adaptação, através de uma seqüência de estágios intermediários, reunindo, de modo sucessivo, elementos que só parecem ter significado como partes de um produto final". *Darwin e os grandes enigmas da vida*, pp. 35 e 98.



mutações que sofre. Pressupondo-se a hipótese básica da explicação funcionalista, uma mutação só é aceita se produzir um efeito benéfico nos organismos descendentes daquele que a sofreu.

Se as mutações herdadas determinarem a aparição, na descendência, de traços morfológicos, fisiológicos e comportamentais superiores, isto é, variações mais favoráveis ao ambiente, ela sobreviverá e se propagará. A superioridade não significa um aumento do grau de perfeição, mas somente que "os organismos ficam melhor adaptados a seus ambientes".<sup>64</sup> O plural aqui, "ambientes", não contradiz a hipótese inicial de um ambiente paramétrico. Isto porque esta hipótese é compatível com a idéia de um ambiente cambiante: o que muda é o ambiente imediato e local, isto é, os estímulos que a cada momento "impressionam" os órgãos dos sentidos dos organismos.

Cada organismo mutante descendente, de uma dada geração, possui, como um dos resultados da mutação aceita, outros critérios para selecionar novas mutações; quer dizer, alteram-se as combinações de estímulos que produzirão mutações resultantes numa superioridade da sua capacidade reprodutiva. Dizendo de outro modo, ele pode ser beneficiado ou prejudicado por *inputs* diferentes dos que antes produziam tais efeitos. Dito brevemente, a "escolha" das mutações é dirigida pelo mecanismo de *seleção natural*: transmitem-se aquelas mutações presentes nos organismos com maior *aptidão*, isto é, os organismos que, por possuírem um *design* aperfeiçoado às novas condições do meio ambiente imediato, duram mais e por isso aumentam a "representatividade dos seus genes nas gerações futuras".<sup>65</sup>

Descrito como uma máquina, cada organismo funciona do seguinte modo: num dado momento do tempo, de acordo com o critério de seleção operante, o mecanismo diz sim ou não a cada *input* recebido. Se em algum ponto da descendência, um organismo alcançar um estado no qual diz não a todos e cada um dos muitos *inputs* possíveis (do conjunto finito de elementos no seu ambiente), então dizemos que alcançou um estado de *máximo global*.<sup>66</sup> Isto quer dizer que, quanto à característica determinada por aquela variação genética particular, a aptidão do organismo encontra-se, relativamente aos demais da população, *maximizada*.

<sup>64</sup> Stephen Jay Gould, *op. cit.*, p. 2.

<sup>65</sup> *Ibid.*, p. 2. Cf. também a discussão do conceito de aptidão, pp. 34-37.

<sup>66</sup> Cf. Elster, *Ulysses and the sirens*, pp. 4-9.

Produz-se desta maneira uma explicação de que a sobrevivência de organismos com um dado fenótipo (formas, fisiologia e comportamento específicos) não é uma mera coincidência, ou obra do acaso. Nesta explicação, a sobrevivência é resultante basicamente de dois mecanismos: *variação* (da qual a mutação é um tipo) e *seleção*. A seleção é uma forma de avaliação da aptidão dos organismos para sobreviver em meios ambientes cambiantes. A aptidão dos organismos é o que nos diz por que os organismos melhores, uma vez que são criados, terminam sobrevivendo.<sup>67</sup>

### 2.7. Uma explicação funcionalista do mecanismo de autopreservação.

Vejam agora como poderia funcionar o modelo biológico quando aplicado ao mecanismo produtor de prazer segundo uma explicação funcionalista coerente com os pressupostos da psicologia materialista. Gostaria, em primeiro lugar, de observar que não tenho qualquer pretensão de estar propriamente apresentando o pensamento de Hobbes; portanto, creio com isso conseguir evitar ser acusado de estar produzindo "uma variante fantasiosa do que [Hobbes] teria podido dizer" ou, menos gravemente, realizando "manobras analíticas ... particularmente delicadas" ao inserir conceitos ou suposições "totalmente estranhas ao argumento hobbesiano".<sup>68</sup> A discussão que estou fazendo não é um número de acrobacia mental, mas somente um artifício de exposição para ilustrar algumas conseqüências teóricas indesejáveis que resultariam da eliminação de qualquer traço de intencionalidade como elemento explicativo da operação do motivo fundamental - segundo a teoria de Hobbes - do comportamento humano.

Existe uma passagem crucial do *Leviathan* que dá uma base razoável para o uso da analogia que proponho:

*"Dado que a constituição do corpo de um homem se encontra em constante modificação, é impossível que as mesmas coisas nele provoquem sempre os mesmos apetites e aversões ..."*<sup>69</sup>

<sup>67</sup> Cf. Herbert Simon, *Reason in human affairs*, pp. 50-65 na edição portuguesa.

<sup>68</sup> Cf. advertência feita pelo professor Balthazar Barbosa Filho; *Condições da autoridade e autorização em Hobbes*, pp. 65 e 68.

<sup>69</sup> 6, 24.

Assim, podemos primeiramente conceber as "modificações" na constituição do corpo de um homem como o correlato, no modelo que estou apresentando, das mutações no modelo da seleção natural. O corpo de um homem pode ser concebido como uma máquina que recebe, sucessivamente, estímulos de diferentes objetos situados no ambiente (vamos, por hipótese, postular a igual probabilidade de qualquer um dos objetos estimular o corpo num dado momento, o que garante que as modificações se dão *ao acaso*).

Os objetos estimulam aleatoriamente o corpo, produzindo pequenas modificações. Uma modificação é uma alteração do movimento vital ou *esforço* (*endeavour*), que é produzido como efeito da ação de um objeto externo nos órgãos dos sentidos. O *esforço* pode gerar um movimento externo de alguma parte do corpo em direção ao objeto que o causa ou um movimento de afastamento deste objeto. Independentemente da direção do movimento voluntário (i. e., movimento externo do corpo) iniciado com o *esforço*, uma modificação não pode ter deixado de ocorrer, produzindo ou não um efeito benéfico no corpo. O benefício é medido pelo aumento do movimento vital, cuja conseqüência capital é o fato de este homem sobreviver por mais tempo, aumentar a duração da sua vida nos momentos futuros mais próximos. O seu corpo terá condições de durar por um maior período de tempo. Por sua parte, o dano é medido pela diminuição do movimento vital, com a conseqüente diminuição do tempo futuro próximo de sobrevivência do homem.<sup>70</sup>

Podemos supor que o corpo dispõe, num momento qualquer do tempo, de critérios bem definidos para *aceitar* ou *rejeitar* aquela modificação inicial. Enquadrando este modelo como explicação funcionalista, é preciso igualmente supor que uma modificação inicial é funcionalmente aceita quando tornar o organismo mais apto para

---

<sup>70</sup> Gauthier faz um comentário semelhante; porém, num contexto argumentativo cujo objetivo não é construir uma explicação de como a ação humana - o que Hobbes denomina por "movimento voluntário" - é, em última instância, direcionada à (tem como fim a) manutenção do movimento vital (cf. *Leviathan*, 6, 23 e *De homine* 7; 1, 2 e 5): "num dado momento do tempo, uma pessoa exhibe um certo nível de movimento vital. Um aumento deste nível eleva a qualidade e a provável duração da sua vida; uma diminuição reduz a qualidade e encurta a provável duração" (*Moral dealing*, p. 131). Ao contrário, penso estar oferecendo uma explicação no sentido de uma perspectiva exibindo uma visão hipotética coerente de como é possível o fato em questão, ou melhor, exibindo hipóteses das quais podemos derivar dedutivamente a proposição que descreve tal fato. Ver o conceito de *explicação filosófica* exposto por Robert Nozick, *Philosophical explanations*, pp. 8-18.

perdurar durante um maior período de tempo (o que valia antes para organismos descendentes, vale agora para o mesmo corpo num tempo futuro próximo).

Uma maior *aptidão* na capacidade de aumentar o movimento vital significa diretamente poder alcançar, num tempo próximo, um patamar maior de prazer; mas, principalmente, significa um aumento de sua capacidade de beneficiar-se mais de alguns futuros estímulos (e prejudicar-se menos com aqueles estímulos que efetuam a diminuição do movimento vital). Tal significado da *aptidão* é derivado do fato de que uma dada modificação, sofrida num certo momento, também funciona como um novo critério de seleção das modificações futuras.<sup>71</sup>

Aceita esta hipótese adicional, pode-se supor unicamente a possibilidade de, *passo a passo*, o corpo chegar a alcançar estados nos quais um dado estímulo produz diretamente um máximo de prazer (ou, indiretamente, ao produzir um mínimo de desprazer). Tal estado corresponde ao *máximo global* no modelo explicativo agora em questão. Neste modelo, a autopreservação é explicada de modo análogo à sobrevivência dos mais aptos na teoria da evolução. A autopreservação é alcançada por meio de um mecanismo de ascensão gradual e contínua da capacidade do corpo do homem de manter seu movimento vital,<sup>72</sup> um tal mecanismo só pode operar por meio de transições contínuas - e indiretas - entre *máximos locais*, até alcançar um máximo local que coincida com o global. Razões mais convincentes a favor desta última afirmação serão apresentadas no apêndice III, por uma consideração matemática elementar, o que permitirá também uma

<sup>71</sup> Pode-se pensar nesta maior aptidão ou capacidade de sobrevivência como uma variável determinada diretamente por uma melhor memória dos objetos que afetaram os sentidos, pelo menos um tipo de memorização que signifique "uma *sensação diminuída*, ... quando queremos com isso exprimir a diminuição e significar que a sensação é evanescente, antiga e passada, ... pois a contínua mudança no corpo do homem destrói com o tempo as partes que foram agitadas na sensação, de tal modo que a distância no tempo e no espaço têm ambas o mesmo efeito em nós." *Leviathan*, 2, 5. No interior do esquema conceitual da psicologia materialista, atribuir ao homem algum desejo (entendido como um tipo de apetite acompanhado da consciência da meta a ser buscada; cf. definição de Spinoza, *Ética*, parte III, proposição ix, escólio) e algum conhecimento do que é bom para a sua conservação é atribuir algo mais do que o homem de fato possui. Cf. *Leviathan*, 2, 4.

<sup>72</sup> A expressão latina *Natura non facit saltum* ("A natureza não dá saltos") é usada por Darwin no tratado *The origin of species* (1859) e sinaliza o papel do "princípio da continuidade" como fundamental para as explicações sobre a evolução na teoria darwiniana: "Se se puder demonstrar que qualquer órgão complexo [como por exemplo, o olho] tenha podido chegar a existir sem que ele tivesse se formado por meio de modificações numerosas, sucessivas e diminutas, minha teoria estaria absolutamente destruída". Citado por Eduardo Giannetti da Fonseca in *Vícios privados, benefícios públicos?*, p. 228.

maior nitidez sobre pontos contidos nesta e na próxima seção. Para tanto, usarei da distinção conceitual entre *máximo local* e *global* tal como é feita no domínio da teoria das *funções*.

### 2.8. Limitações da explicação funcionalista.

Se o exercício de argumentação feito até aqui é válido, deve servir para fazer com que diferentes passagens dos textos de Hobbes possam ingressar como partes de um mesmo esquema explicativo. Assim, o parágrafo anterior pode ser tomado como um comentário do trecho abaixo:

*"Aceitando que todo prazer é apetite, e pressupõe um novo fim, não pode haver satisfação a não ser na continuidade. Portanto, não devemos nos maravilhar quando constatamos que, na medida que os homens alcançam mais riqueza, honra, ou outro poder, o seu apetite cresce continuamente mais e mais, e quando eles chegam ao mais alto grau de algum tipo de poder, eles perseguem algum outro."*<sup>73</sup>

Ainda assim, resta um sério problema, insolúvel na explicação funcionalista. Visto como um tipo de máquina, o corpo humano não é capaz de realizar a chamada maximização global, algo que é típico da *escolha humana* e que envolve basicamente a capacidade de utilizar dois tipos básicos de mecanismos: *esperar* por uma modificação melhor que aquela que está sendo imediatamente oferecida e adotar *estratégias indiretas*, isto é, aceitar uma modificação que resulta ser imediatamente prejudicial para posteriormente poder ter acesso a outra modificação que é, relativamente à perda causada pela anterior, muito mais benéfica. Como sugere Jon Elster, comentando metaforicamente as limitações características das máquinas capazes apenas de efetuar *máximos locais*:

*"Como um escalador que sobe passo-a-passo, o organismo mantém o seu olho míope fixo no chão, e é incapaz de levar em conta o que acontece atrás da colina seguinte"*<sup>74</sup>

<sup>73</sup> *De homine*, 7, 6. Tradução minha.

<sup>74</sup> *Ulysses and the sirens*, p. 9.

Hobbes parece não ter problema em aceitar o automatismo do mecanismo de ascensão gradual e contínua da capacidade, não só do corpo humano, mas da maioria das coisas corporais, de manter seu movimento vital:

*"Faz parte da natureza de quase todas as coisas corporais que, sendo repetidamente movidas de uma única e mesma maneira, recebem continuamente uma cada vez maior facilidade e aptidão para o mesmo movimento".<sup>75</sup>*

Portanto, Hobbes sabe, ao que tudo indica, que tem de enfrentar o problema em questão, e que a dificuldade é gerada necessariamente sempre que uma teoria limitada como a "psicologia materialista" for usada como base de uma teoria da motivação. Isto tem uma consequência importante: ao contrário do que sustenta Jean Hampton, a "psicologia materialista" não pode, por si só, funcionar como "base fundacional" da teoria do valor que Hobbes quer construir.

A solução, como veremos, passa por um outro tipo de interpretação do papel da autopreservação, agora como constituinte de um comportamento descrito como intencional (envolvendo desejos, crenças, metas, expectativas, capacidade de cálculo e de antevisão das consequências de cursos de ação possíveis, etc.), o que exigirá a introdução da concepção de homem como um tipo de máquina globalmente maximizadora, com a capacidade de fazer escolhas aplicando preceitos constitutivos da *racionalidade*. Uma tal solução pressupõe também a introdução de sofisticações teóricas, que Hobbes mesmo lançará mão, no interior da teoria da racionalidade. Isto é verificável, por enquanto apenas intuitivamente, na comparação simplificada de uma passagem do livro de Herbert Simon antes citado, no qual o autor contrasta um traço unicamente humano - o chamado comportamento racional estratégico - com a "miopia da evolução", e uma passagem na qual Hobbes indica que é próprio da racionalidade a habilidade de conceber estados de coisas futuros, isto é, não dados imediatamente aos sentidos:

---

<sup>75</sup> *De homine*, 5, 14.

"Se planejarmos uma ação através de uma análise consciente e racional e formos suficientemente espertos, podemos antever um determinado período de tempo - pelo menos um período pequeno - e conceber as possíveis conseqüências da ação durante esse mesmo período. Em princípio não existem limites fixos para os horizontes temporais que regem o cálculo das conseqüências da ação. Uma tal capacidade de antevisão contrasta vivamente com os mecanismos da evolução biológica, os quais fornecem apenas um tipo de racionalidade muito míope. A aptidão tende a selecionar aqueles organismos que asseguram uma vantagem imediata a curto prazo. Qualquer que seja o seu ponto de partida, o organismo ascenderá na escala local de aptidão."<sup>76</sup>

"Os homens portanto permanecem no estado de guerra, enquanto medirem diversamente o bem e o mal, segundo a diversidade dos apetites que neles dominam. E não há nenhum sequer que deixe de reconhecer que este estado é mau, e que, conseqüentemente, a paz é uma coisa boa. Aqueles então que não podem colocar-se de acordo sobre o que seja um bem presente, podem-no quanto a um bem que está por vir. Isto é um efeito do cálculo, porque as coisas presentes caem sob os sentidos, mas as futuras só podem ser concebidas pelo raciocínio. De maneira que a razão nos dita que a paz é uma coisa desejável."<sup>77</sup>

Como outra conseqüência, seria igualmente insuficiente dizer, de modo dogmático, que o ser humano deseja *intrinsecamente* tudo o que aumenta o movimento interno das partes do seu corpo, i.e., busca tudo o que proporciona um aumento do seu movimento vital. Tal desejo seria apenas uma *propriedade intrínseca* a mais dos seres humanos, do mesmo modo que o é a propriedade de possuir coração. São ambas intrínsecas porque os seres humanos as têm em virtude somente de serem tal tipo de criatura:<sup>78</sup> o desejo de autopreservação, assim interpretado, pertenceria a uma categoria mais ampla:

<sup>76</sup> *Reason in human affairs*; p. 82 na ed. portuguesa. A evolução biológica depende de um "tipo básico de inteligência" contrastante com a "essência da inteligência humana":

"Somos, como Konrad Lorenz afirmou, 'especialistas em não-especialização'. Nós não nos comportamos qual máquinas, com simples comandos de sim/não, invariavelmente acionados por fragmentos de informações presentes em nossas vizinhanças imediatas. (...) Todavia, esse modelo inflexível representa o estilo de inteligência adotado com grande sucesso pela maioria dos outros animais. As decisões tomadas pelos animais são insofismáveis sim(s) e não(s), acionados por sinais determinados e não por sutis escolhas baseadas na avaliação de uma complexa *gestalt*." Stephen J. Gould, *Hen's teeth and horse's toes*; pp. 51-52 na trad. brasileira.

<sup>77</sup> *De cive*, III, XXXI, 120.

<sup>78</sup> Isto é sustentado por Jean Hampton, *op. cit.*, pp. 8 e 14.

"Dos apetites e aversões, alguns nascem com o homem, como o apetite pela comida, o apetite de excreção e exoneração (que podem também, e mais propriamente, ser chamados aversões, em relação a algo que se sente dentro do corpo) e alguns outros apetites, mas não muitos"<sup>79</sup>

### 2.9. Ações não podem ser explicadas pela "psicologia materialista".

A insuficiência de uma "psicologia materialista" diz respeito não só à função do desejo de autopreservação, mas, antes disso, à explicação das ações ou movimentos voluntários. Nas explicações ordinárias das ações, sempre incluímos algum "desejo que compreende, de maneira essencial, o futuro - querendo dizer com 'o futuro' 'algo causalmente subsequente à ação'.<sup>80</sup> Uma maneira contemporânea de enunciar esta tese é dizer que as condições de verdade de enunciados do tipo "A queria (pretendia, tinha a intenção de) obter  $x$  por meio de uma certa ação" são (i) "A esperava que tal ação o levaria a obter  $x$ " e (ii) "A expectativa de A obter  $x$  explica por que A realizou tal ação".<sup>81</sup> Assim, se quisermos explicar o movimento de alguma parte do corpo de A como uma ação, não basta dizer que tal movimento foi causado por algum estado *intrínseco* de A (i. e., alguma modificação do seu *movimento vital*); é preciso fazer referência a pensamentos, crenças e desejos de A relativos a estados de coisas futuros, "objetos" (termo genérico de Hobbes para coisas e estados de coisas) ausentes, quer dizer, que não estão causando *sensações* ao pressionarem seus órgãos dos sentidos.

E isto é precisamente o que Hobbes está pretendendo ao afirmar que "[todos os movimentos voluntários] dependem sempre de um pensamento anterior de *como, onde e o que*".<sup>82</sup> Portanto, a imaginação (entendida como faculdade de imaginar, de ter cadeias de imaginações, pensamentos ou concepções - *discurso mental*),<sup>83</sup> prossegue Hobbes, "é a primeira origem interna de todos os movimentos voluntários". Quando Hobbes afirma que as coisas "futuras só podem ser concebidas pelo raciocínio", ele quer dizer que só podemos conceber coisas que ainda "estão por vir", só podemos antevê-las, por meio da imaginação.

<sup>79</sup> *Leviathan*, 6, 24.

<sup>80</sup> Jonathan Bennett, *A study of Spinoza's ethics*; p. 229 na ed. mexicana.

<sup>81</sup> Cf. Bennett, *op. cit.*, p. 232.

<sup>82</sup> *Leviathan*, 6, 23.

<sup>83</sup> Ver *Leviathan* 2, 8 e 3, 8.

! a imaginação  
que calcula?



Dadas as restrições que condicionam o uso do raciocínio entendido como cálculo (ver cap. 4), é preciso primeiro entreter no espírito os resultados ou conseqüências que concebemos como acontecimentos causados por nossas ações; e depois, atribuir a cada uma destas "ficções do espírito" - cujas existências futuras são esperadas no presente com maior ou menor expectativa - um determinado valor.<sup>84</sup>

A imaginação, entendida genericamente como "capacidade de antevisão", é descrita - marcadamente sem maiores contrastes com a apresentação hobbesiana desta capacidade - pela filósofa inglesa contemporânea Mary Warnock da seguinte forma:

*"Uma pessoa totalmente sem imaginação - se fosse possível existir tal pessoa - seria totalmente limitada pelas restrições da experiência imediata ... Quando falamos que a imaginação lida com aquilo que não é, queremos dizer que é por meio dessa faculdade que pensamos nas coisas na sua ausência. E se podemos fazer isto, podemos também concebê-las como não são, ou ainda não são, mas como podem vir a ser"*<sup>85</sup>

Segundo Hobbes, as condições para se dizer que um certo movimento é voluntário (i. e., uma ação) são:

1. que ele seja de alguém que espera que exista algo no futuro produzido como resultado do próprio movimento (para Hobbes "ter apetite ou medo de *x*" significa respectivamente "ter expectativa de um *prazer futuro - fruição de x*, um *fim atingido - ou ter expectativa de um *desprazer futuro - aversão de x*");*

1.1. para que alguém possa ter *expectativa de bem*, deve ser capaz de conceber e por isso esperar, agora, que seu movimento tenha o *poder* de produzir, no futuro, algo que deseja agora; da mesma forma, para que alguém possa ter *expectativa de mal* deve ser capaz de conceber e por isso esperar, agora, que seu movimento tenha o *poder* de produzir, no futuro, algo de que tem medo agora;

1.1.1. para que alguém possa ter *expectativa de bem* ou *de mal*, deve ser capaz de *recordar* que, até o momento, algo que deseja ou de que tem medo tem sido produzido por meio de movimentos similares (ou, deve ser capaz de conceber um *poder* semelhante passado e um ato futuro de um *poder* semelhante).<sup>86</sup>

Repassando estas considerações sobre os elementos envolvidos nos pensamentos que antecedem todos os movimentos voluntários, poderíamos incluir também o *tempo* entre os fatores que estruturam as operações da capacidade de imaginar. Assim,

<sup>84</sup> Ver *Leviathan*, 3, 10.

<sup>85</sup> *The uses of philosophy*; pp. 189-190 na ed. brasileira.

<sup>86</sup> Cf. *De homine*, 8, 3 e 12, 2.

está justificado o seguinte acréscimo à afirmação de Hobbes: os movimentos voluntários *também* dependem de um pensamento anterior de *quando*.

### 2.10. A intencionalidade das paixões.

Deslocando o foco da discussão feita até aqui, é possível considerar que as afirmações acima numeradas tornam a teoria da motivação hobbesiana não inteiramente derivável da "psicologia materialista". Bennett apontou com precisão a consequência radicalmente inaceitável da adoção de uma teoria da motivação de cunho fisicalista:

*"O que importa é que Hobbes diga que um apetite ou desejo de x deve ser causado por x. Isso me permite dizer que algumas das consequências das ações próprias não são desejadas, mesmo se são esperadas; meu movimento em direção à maçã foi também em direção ao gusano ou foi causado somente pela maçã e, deste modo, foi um desejo somente da maçã. Por esta vantagem [Bennett está comparando-a com a teoria de Spinoza], a teoria de Hobbes paga um preço elevado; não se considerará que eu tenha um apetite de x a menos que x tenha atuado e, por isso, ela nada tem a dizer sobre o fato de eu regar a árvore da qual quero obter maçãs no próximo outono"*<sup>87</sup>

Na descrição do comportamento humano, é preciso então introduzir o traço de intencionalidade, alteração esta que nos conduzirá, no final, entre outros resultados importantes, a uma interpretação mais adequada do desejo de autopreservação como elemento central na explicação das ações.

Segundo Hobbes, o início de qualquer ação (entendida como *comportamento intencional*) é uma paixão. Podemos traduzir esta concepção dizendo que a base motivacional de toda ação é um desejo; este é "o único elemento independente, ao qual todos os outros são subservientes".<sup>88</sup>

Com o termo *paixão*, próprio da filosofia moral moderna, Hobbes nomeava certos estados da "imaginação" (atualmente diríamos, alguns tipos de "estados mentais")

<sup>87</sup> *Op. cit.*, p. 232.

<sup>88</sup> Elster, *Nuts and bolts for the social sciences*, p. 30. Esta concepção é geralmente associada a Hume. A tese "humeana" pode ser enunciada como "toda ação necessariamente inclui um desejo". Gaus, *op. cit.*, p. 85.

cujo termo genérico hoje é *desejo* (incluem-se como tipos particulares de desejos os *apetites* e as *emoções*). Aqui, é preciso fazer algumas considerações sobre a natureza das paixões. Genericamente, pode-se dizer que as paixões são "sensações que agitam a mente"<sup>89</sup> Desde o ponto de vista da "psicologia materialista", como já vimos, as paixões são apenas um tipo especial de causa inicial dos movimentos corporais: *esforço* (*endeavor*). Ele é causado pela ação dos objetos externos e causa, por sua vez, o movimento voluntário que em parte descreve as ações.

No entanto, tal esforço não é percebido pela mente como um movimento, mas como um tipo de sensação que pode ser descrita como uma *perturbação da mente* (*trouble of mind*).<sup>90</sup> O problema é que o termo "perturbação" é genérico e, por esta razão, impreciso. Se se entende este termo como significando um "evento mental", ele é genérico porque, podemos dizer, se aplica indistintamente a "tudo aquilo que está de tal modo em nós que somos imediatamente conscientes".<sup>91</sup> Neste sentido, a sensação de uma dor ou de uma cor são igualmente "perturbações" da mente.

Aceitando a posição de Hobbes de que as paixões são representadas na mente como sensações ou afecções, não como movimentos corporais (tal como assinalou Descartes, respondendo a uma objeção de Hobbes, uma representação ou imagem de uma coisa não significa uma cópia ou reprodução da coisa produzida pela imaginação, mas a expressão mental dela),<sup>92</sup> é preciso encontrar algum traço específico para distingui-las de outras *maneiras de pensar*. Procurando usar uma terminologia cartesiana, pode-se dizer que as paixões referem-se a certos *atos de querer*, por oposição a *atos de pensar* que envolvem somente *ter percepções*, sem qualquer operação da vontade associada (por exemplo, sentir, e *imaginar*, o que aqui significa ter uma cadeia de pensamentos que não é regulada por algum desejo ou desígnio).<sup>93</sup>

Pensamentos, neste sentido, i. e., incluindo formas de *cognição* e *volição*, são fenômenos *intencionais*, isto é, são *sobre* alguma coisa, *de* alguma coisa. No caso dos *atos*

<sup>89</sup> Spinoza, *Tractatus politicus*; citado por E. Giannetti da Fonseca, *Beliefs in action*, p. 90.

<sup>90</sup> Cf. *Leviathan*, 6, 23-25.

<sup>91</sup> Definição cartesiana de *pensamento*; *Exposição geométrica*, 1ª definição.

<sup>92</sup> Cf. Raul Landim Filho, *Análise da noção de objeto e representação em uma filosofia da consciência e em uma filosofia lógico-lingüística*, pp. 495-496.

<sup>93</sup> Em *Evidência e verdade no sistema cartesiano*, pp. 47-53, o professor Raul Landim apresenta a distinção cartesiana entre os pensamentos que são "atos perceptivos" e os que são "atos da vontade" - estes últimos incluindo, além de desejar e ter aversão, crer, duvidar, afirmar e negar.

*de querer* em questão, expressáveis pela forma geral "o desejo de  $x$ ", existe um tipo especial de objeto cujo nome pode preencher o lugar vazio indicado pela variável: um *fim*. O objeto do desejo é sempre um fim, uma meta, algo que se quer alcançar, obter, realizar. Podemos acrescentar uma outra definição: se alguém deseja alguma coisa, dizemos que tem a intenção de (tem como propósito, como meta) alcançar, obter, realizar tal coisa. Os conceitos de desejo e intenção são conceitos teleológicos, porque na sua análise é usado o conceito de fim ou meta. Assim, poderíamos definir implicitamente uma paixão dizendo que a expressão "ter uma paixão" significa visar (alcançar, obter, realizar) alguma coisa.<sup>94</sup>

Resumindo o ponto: segundo uma das descrições acima, as paixões são diferentes sensações ou percepções mentais - diferentes "estados de agitação mental" - de mudanças corporais, ou melhor, de alterações no movimento vital em diferentes partes do corpo, que, por sua vez, foram causadas pela percepção - por meio dos órgãos dos sentidos - de um objeto. Para sermos coerentes com esta descrição, podemos dizer que existem basicamente dois "estados de agitação mental" ou, para usar os termos de Hobbes, duas "aparências ou sensações" possíveis das modificações do movimento vital: prazer (ou deleite) e desprazer ou dor (ou desagrado). Estas "sensações" correspondem respectivamente ao aumento e à diminuição do movimento vital, e geram o que Hobbes chama de *esforço*, um tipo de força ou impulso que causa o início, "no interior do corpo do homem", das ações.

No entanto, se as paixões não fossem senão meros impulsos percebidos pela mente como sensações agradáveis ou desagradáveis, então todas as ações humanas seriam descritíveis como movimentos cegos ou sem propósito, isto é, como derivadas de propensões para "fazer coisas que não estão relacionadas com qualquer expectativa de satisfação, prazer, ou qualquer outra crença explicativa."<sup>95</sup> Conforme outra descrição, quando dizemos que alguém "tem uma paixão" (ou que está num "estado emocional  $E$ "), ainda que com isto queiramos descrever uma certa sensação ou afecção que tal pessoa está experimentando, estamos também querendo informar *sobre o que é afinal o estado emocional  $E$* . Para tanto, a análise das *emoções* ou *paixões* deve poder apresentá-las como algo que se define por uma *afecção* associada a uma *crença*. Esta definição é fundamental na concepção hobbesiana da natureza das *paixões*.

<sup>94</sup> Cf. Anscombe, *Modern moral philosophy*, p. 2.

<sup>95</sup> Gaus, *op. cit.*, p. 91.

*Afecções* são tipos de *sensações* que Hobbes chama de "*paixões simples*". Há, em última instância, dois tipos irredutíveis de "paixões simples": o *prazer* (ou *deleite*) e o *desprazer* (ou *desagrado*). Cada um deles, por sua vez, pode derivar da "sensação de um objeto presente" - são os chamados *prazeres dos sentidos* - ou da "expectativa provocada pela previsão do fim ou consequência das coisas" - são os *prazeres do espírito*.<sup>96</sup>

A *diversidade das paixões* depende de dois componentes: da *diversidade da sensação* - pois as coisas nos afetam de diferentes modos, conforme as constituições do corpo humano - e da *diversidade da opinião*. Este último componente indica não somente que as paixões são causadas por juízos sobre os seus objetos, mas também que são *conceitualmente dependentes* das crenças expressas por tais juízos - como diz Hobbes, "as paixões simples ... recebem nomes diversos conforme a maneira como são consideradas".<sup>97</sup>

Vejamos dois exemplos:

- (1) *A* tem *esperança* se e somente se:
  - (i) *A* deseja *x* ; e
  - (ii) *A* crê ser possível obter *x*.<sup>98</sup>

<sup>96</sup> Ver *Leviathan*, 6, 25. Que as emoções - as "paixões", no vocabulário do séc. XVII - tenham um elemento sensorial não é um ponto disputado mesmo nas atuais teorias psicológicas. Além disso, tal como Hobbes aponta, as *sensações* das paixões consistem em percepções internas de movimentos em alguma parte do corpo; no entanto, qualquer mudança corporal só pode ser percebida internamente se resultar da transmissão, por meio dos nervos, do movimento até o cérebro. Ver *Leviathan*, 2, 6. Segundo entendo, Hobbes descartaria como insuficiente o que hoje conhecemos como "teorias periféricas das emoções". A mais importante delas, sustenta Daniel Robinson, foi a desenvolvida por William James (tenho dúvidas acerca da correção de sua descrição da teoria de James) e C. G. Lange. Conforme esta teoria, "uma emoção *sentida* é simplesmente a consequência [de] alterações periféricas: 'vemos um urso, corremos, e então sentimos medo'. Em outras palavras, o estímulo (o urso) ativa o comportamento simpático do sistema nervoso autônomo, alterando assim uma série de eventos biológicos (viscerais); a *soma* destas alterações é a emoção do medo". *The enlightened machine: an analytical introduction to neuropsychology*; p. 140 na trad. brasileira. Gaus apresenta um dos experimentos decisivos para a não aceitação do que denomina de "*somatic-viscera theories of emotion*" (i. e., as "teorias periféricas"); ver *Value and justification*, pp. 33-34.

<sup>97</sup> *Leviathan*, 6, 25. Uma análise precisa da *intencionalidade* das paixões incluiria a *crença* e o *desejo* como "estados intencionais básicos"; o que excluiria a possibilidade da redução de um desejo a uma mera *afecção*, já que deve sempre possível responder a perguntas como "A que se refere tal desejo?", "Em que consiste aquilo que você deseja?". As paixões podem ser representadas na mente apenas parcialmente como sensações ou afecções, pois, enquanto pertencentes a uma *categoria tradicional da Volição*, não podem ser reduzidas a estados mentais *não-direcionados para* objetos e estados de coisas no mundo, tal como é o caso de um "sentimento súbito de exaltação". Estas observações estão baseadas no cap. 1 do livro *Intencionalidade*, de John Searle.

<sup>98</sup> "O *apetite*, ligado à crença de conseguir, chama-se *esperança*". *Idem*. *Esperar* é, na verdade, sempre *esperar que p*, cuja análise pode incluir um *grau de crença* de que *p* de fato se verifique no futuro ou a "*completa incerteza*" quanto à ocorrência futura de *p*, incerteza que poderia ser traduzida pela conjunção *não crê que p e não crê que não-p*; neste último caso, alguém *espera que p* porque

(2) *A* sente *medo* se e somente se:

(i) *A* tem aversão a *x*; e

(ii) *A* crê que *x* é uma ameaça, por poder causar-lhe algum dano.<sup>99</sup>

Esta análise - fiel ao texto de Hobbes- tem uma vantagem capital: poder explicar como é possível que as paixões sejam *inadequadas* dadas certas circunstâncias. Para isso, basta pensar que as crenças às quais se associam não são, dadas as evidências disponíveis na situação (o que inclui também a coerência com outras crenças sustentadas pelo indivíduo), crenças racionais. Como consequência, a análise dos tipos de erros que levam os agentes a fazerem escolhas irracionais torna-se mais fina. (Este ponto será melhor apreciado na seção 4.11.)

---

*deseja que p seja o caso e crê ser possível que p seja o caso, ainda que não tenha nenhuma expectativa de que p. Ver Searle, op. cit. No cap. 4 desta dissertação, será adotada a concepção de racionalidade segundo a qual, sob condições de incerteza, é razoável para um agente fazer suas escolhas com base nas expectativas que tem da ocorrência dos eventos que se seguem como consequências de ações cujas alternativas tem a seu alcance.*

<sup>99</sup>”A *aversão*, ligada à crença de *dano* proveniente do objeto, chama-se *medo*”. *Idem*. A edição brasileira (“*Os pensadores*”) contém um erro nesta passagem. Lê-se nela, em vez de “aversão”, o termo “opinião”.

### 3. TEORIA DO VALOR

*"Uma coisa só tem valor para quem a deseja"*

John Steinbeck, *The winter of our discontent*

*"Todo valor é uma grandeza comparativa e, até mesmo, apresenta-se necessariamente em dupla relação: pois primeiro é relativo já que é para alguém e, segundo, é comparativo pois está em comparação com alguma outra coisa, de acordo com a qual é avaliado. Retirado destas duas relações, o conceito de valor perde todo o sentido e o significado".*

Schopenhauer, *Über das Fundament der Moral*

#### 3.1. A natureza dos juízos de valor.

A concepção fundamental da teoria do valor de Hobbes está esboçada na seguinte afirmação:

*"É necessário então saber que os termos bom e mau são nomes impostos às coisas a fim de testemunhar o desejo ou a aversão daqueles que lhes dão este título."<sup>100</sup>*

São duas as teses conexas aí afirmadas:

1. Quando alguém diz que "x é bom", está querendo simplesmente dizer que "deseja x" (isto está indicado na expressão "os termos [são usados] a fim de testemunhar...");

2. Da análise do significado de "bom" exposta em 1, está implicado que "bom" e "mau" são nomes atribuídos aos objetos; quer dizer, não são propriedades misteriosas dos objetos, mas apenas "etiquetas" coladas aos objetos para indicar se eles causam desejo ou aversão numa determinada pessoa. Dito de outra forma, por meio do juízo "x é bom" não se atribui uma propriedade a um objeto. Hobbes está sustentando que os valores *bom/mau* são *relativos* àqueles que avaliam; por isso, "x é bom" deve ser analisado como significando "x tem a relação R com uma pessoa y".

A relação R pode ser basicamente analisada como: "... é útil para ..." (com  $\sim R =$  "... é prejudicial para ...") e "... é prazeroso para ..." (com  $\sim R =$  "... é doloroso para ...").<sup>101</sup>

<sup>100</sup> *De cive*, III, XXI, 119.

<sup>101</sup> Cf. *Leviathan*, 6, 24-25.

Estas duas teses aparecem de modo mais explícito na seguinte passagem do *Leviathan*:

Correspondendo a 1: "Mas seja qual for o objeto do apetite ou desejo de qualquer homem, esse objeto é aquele a que cada um chama *bom*; ao objeto de seu ódio e aversão chama *mau*, e ao de seu desprezo chama *vil e indigno*.

E a 2: "Pois as palavras 'bom', 'mau' e 'desprezível' são sempre usadas em relação à pessoa que as usa. Não há nada que o seja simples e absolutamente, nem há qualquer regra comum do bem e do mal, que possa ser extraída da natureza dos próprios objetos".<sup>102</sup>

Apresentando tais teses como respostas a problemas meta-éticos (problemas "de segunda ordem" da ética, segundo outra terminologia), pode-se resumi-las da seguinte forma:

1. Juízos de valor são equivalentes a relatos dos sentimentos de quem os profere (*tese sobre o significado dos juízos valorativos*);

2. Juízos de valor são relacionais, isto é, usam como predicado um tipo de relação entre objetos e desejos ou satisfações possíveis de quem os profere (*tese ontológica sobre a referência dos termos valorativos*).

---

<sup>102</sup> 6, 24. No *De homine*, 7, 3 - edição *English works*, tomo IV - encontra-se a seguinte afirmação: "[Os homens] também diferem uns dos outros naquilo que concerne à distinção comum entre o bem e o mal. *Também não existe aqui algo como um bem absoluto, considerado sem relação.*"; grifo meu; tradução de Carlos Alberto Ribeiro de Moura; citado por Luiz Roberto Monzani, *Desejo e prazer na idade moderna*, p. 85.



### 3.2. A continuidade pessoal como maior bem.

Já é possível explicitar uma proposição importante da *teoria do valor* de Hobbes: o que *parece* bom a cada um (o termo "parece" indica explicitamente o caráter intensional do juízo de valor), o que cada um deseja, é o que cada um *julga* ser capaz de lhe proporcionar prazer (aumentar o movimento vital). Deste modo, cada um não pode deixar de desejar - e visar obter - aquilo que julga ser capaz de lhe proporcionar prazer.<sup>103</sup>

Explicitada tal proposição, é possível entender por que o medo da morte é o maior medo. É a morte que traz a pior conseqüência possível para o ser humano, um ser que tem aversão à dor (ausência de prazer): o bloqueio de todo movimento vital, o que significa a aniquilação do indivíduo, quer dizer, a interrupção do que garante a *continuidade pessoal*. Embora Hobbes não use a expressão "continuidade pessoal", pode-se sugerir um significado para ela, a partir de algumas de suas considerações. Cito uma: "... the object of mans desire, is not to enjoy once onely, and for one instant of time; but to assure for ever, the way of his future desire".<sup>104</sup> A passagem sugere que o que cada um mais deseja, o seu *maior bem*, é garantir, no presente, condições para satisfazer *seus* desejos futuros; ainda que não saiba *agora* quais serão eles.

Podemos pensar este desejo como um desejo de (pelo menos) segunda ordem: o objeto deste desejo, de uma pessoa no presente, é a satisfação de outros desejos, desta *mesma* pessoa no futuro. Existem então duas condições para a sua satisfação. Primeiramente, a continuidade do movimento vital, que é o que garante o "*contínuo* progresso do desejo, de um objeto para outro, não sendo a obtenção do primeiro outra coisa senão o caminho para conseguir o segundo".<sup>105</sup> Esta condição é insuficiente porque ela garante somente a continuidade causal da cadeia de desejos, e o interesse de cada um não se limita a conseguir que certos futuros desejos, que no presente ainda não se formaram, sejam de fato *seus* no futuro. Também é do interesse de cada um que, a satisfação dos seus desejos presentes seja um caminho para a satisfação de seus desejos futuros. Para satisfazer

<sup>103</sup> Esta é a interpretação hobbesiana da proposição que, por ser muitas vezes aceita como auto-evidente, figura como um axioma em teorias éticas de diferentes períodos históricos: *todas as pessoas agem visando algum bem*. Urmson analisa, por exemplo, as interpretações socrática e aristotélica desta proposição. Ver *The goals of action* in J. Raz (ed.), *Practical reasoning*, principalmente pp. 131-133.

<sup>104</sup> *Leviathan*, 11, 47.

<sup>105</sup> *Idem*; grifo meu.

esta condição, é preciso que a satisfação dos desejos presentes represente um acúmulo de meios de satisfação dos desejos futuros.

### 3.3. O conceito de poder.

Um meio qualquer de satisfação de desejos é o que Hobbes denomina de *poder*. Assim, o aumento no presente do movimento vital é uma condição direta ou indireta para o aumento de poder. Direta quando pensamos na melhoria de capacidades corporais e intelectuais; indireta quando pensamos na obtenção de instrumentos que auxiliam ou ampliam o uso delas, como por exemplo, saber, honra e riqueza.<sup>106</sup>

A autopreservação é a própria continuidade da satisfação dos desejos que seguirão a formar-se. Como já vimos, a autopreservação é a interpretação hobbesiana de um conceito básico da ética: *felicidade*. "Felicidade" é a tradução usada por Hobbes e, tradicionalmente, por outros autores de língua inglesa do termo grego *eudaimonia*. Hobbes define-a como "o *sucesso contínuo* na obtenção daquelas coisas que de tempos em tempos os homens desejam, quer dizer, o prosperar constante".<sup>107</sup>

Esta não é uma interpretação aristotélica do termo. Seguindo uma observação feita anteriormente, *a felicidade não é um bem absoluto* porque não é algo que possa ser julgado como bom independentemente de ser desejado (todo bem é *relativo*, i. e., tudo o que é um bem o é porque é julgado por alguém como um bem). E igualmente, *a felicidade não é um fim último* porque este suporia a existência de um desejo cuja satisfação em

<sup>106</sup> O conceito de *poder* ocupa uma função explicativa nuclear na *teoria da motivação* hobbesiana: "toda a concepção do futuro é a concepção de um poder capaz de produzir algo; portanto, quem tem a expectativa de um prazer futuro, deve além disso, conceber em si mesmo algum poder pelo qual esse prazer possa ser atingido ... Por este poder entendo o mesmo que as faculdades do corpo e da mente; ... do corpo: nutritiva, geradora, motora; da mente: conhecimento. E além destes, os poderes mais longínquos pelos quais são adquiridos: riquezas, posição de autoridade, amizade ou favor, e boa sorte, que, afinal, nada mais é realmente senão o favor da Omnipotência Divina". *De homine*, 8, 3 e 4.

<sup>107</sup> *Leviathan*, 6, 29.

algum momento do tempo não fosse também um meio para garantir a satisfação de futuros desejos.<sup>108</sup>

O desejo de autopreservação pode ser então definido como o desejo de garantir atualmente, e "para sempre", os meios para a satisfação de seus desejos futuros. Usando a definição de poder antes apresentada, o desejo de autopreservação equivale a um "perpétuo e irrequieto desejo de poder e mais poder" (*Power after power*), que cessa apenas com a morte.<sup>109</sup> Mas o que pode significar para alguém buscar "garantir para sempre os meios de satisfação de seus desejos futuros", ou, simplificando, buscar satisfazer seu desejo de autopreservação? Se a autopreservação é - tal como dizia Aristóteles da *eudaimonia* - um composto de muitos bens, que *não conta como um* dentre eles, ou como hoje denomina-se, um bem ou fim de ordem superior, não é possível por isso "compô-la" de modos diversos? Por exemplo, avaliando as coisas que se busca pelo maior ou menor "deleite sensual" - o prazer dos sentidos -, buscar fazer sempre aquilo cujas conseqüências sejam as mais sensualmente prazerosas? Ou quem sabe, para alguém que quer aparentar ser glorioso (e que se apraz em consegui-lo aparentar), agir buscando vangloriar-se de si mesmo - supondo-se capaz de feitos poderosos e meritórios (heróicos) - e de modo a receber dos outros tantas lisonjas quanto possível? No primeiro caso, ingressam como bens coisas cuja fruição agrada aos sentidos; no segundo, tudo o que serve para aparentar a glória, através do falso reconhecimento do próprio poder e mérito, seja por auto-engano ou por engano dos outros.

---

<sup>108</sup> Cf. *De homine*, 7, 6 e 7. Para uma crítica das impropriedades daquela tradução, bem como uma apresentação das diferenças mais relevantes entre os sentidos aristotélico e moderno do termo, ver especialmente Henry Sidgwick, *The methods of ethics*, pp. 92-93; Bernard Williams, *Ethics and the limits of philosophy*, cap. 3 e J. O. Urmson, *Aristotle's ethics*, cap. 1.

<sup>109</sup> *Leviathan*, 11, 47.

### 3.4. A cobiça expressa uma atitude valorativa em relação ao tempo.

Vamos supor também que tais pessoas, ainda que discordem quanto ao que consideram vantajoso (porque visam coisas diferentes)<sup>110</sup>, coincidem na preocupação em obter, daquilo que desejam, quanto mais possível num tempo próximo. A variável *tempo* é essencial na caracterização deste tipo de pessoa: as conseqüências de uma dada ação perdem rapidamente o seu valor a medida que mais se distanciam no futuro, na prática podendo até mesmo ser completamente ignorada. Esta disposição a agir de modo a ter no presente sempre mais é o que Hobbes denomina de "*greed*", termo que se pode traduzir como *ganância*, *avidez* ou *cobiça*.<sup>111</sup> Ela resulta, de acordo com Hobbes, de um "apetite brutal" (ou também "um desejo desregrado") por vantagens decorrentes das conseqüências mais próximas das ações, apetite que explica a *miopia*<sup>112</sup> com relação às conseqüências mais longínquas, mesmo que sejam bastante desagradáveis.

O problema, tal como o interpreto, não está no fato de que cada ação tem muitas conseqüências às quais não se pode atribuir uma probabilidade (objetiva) de ocorrência (por serem totalmente inesperadas - *novas* conseqüências - ou por serem esperadas apenas com base em palpites, sem um apoio estatístico),<sup>113</sup> mas no fato de tais

<sup>110</sup> É enganoso pensar que, pelo fato de todo ser humano buscar, como "fim último" de sua conduta, a continuidade do prazer, há um acordo sobre *o que é vantajoso*. No entanto, essa "coincidência" de fins é apenas aparente, pois diferentes coisas buscadas (glória, honra, dinheiro, etc.) podem gerar diferentes tipos de prazer, não comensuráveis entre si.

<sup>111</sup> MacIntyre observa que "*greed*" é a melhor tradução para o termo aristotélico *pleonexia*. Este descreve um traço de caráter, quer dizer, uma disposição para agir de um modo específico e por uma razão específica. Hobbes, segundo MacIntyre, "pode ter sido o primeiro escritor de língua inglesa a explicar '*pleonexia*' como sendo 'o desejo de mais do que a sua parte'; porém, acrescenta este autor, esta tradução é enganadora. Pois, para Hobbes, o termo *pleonexia* se aplica unicamente ao tipo de desejo que leva alguém a cometer um tipo de injustiça: desrespeitar a *lei da natureza* segundo à qual ninguém pode deixar de renunciar a qualquer *direito de natureza* e, ao mesmo tempo, exigir que seja renunciado pelos outros, pela razão de não se considerar como igual *por natureza* aos outros. Ver MacIntyre, *Whose justice? Which rationality?*; pp. 125-126 na edição brasileira; e *Leviathan*, 15, 76-77; especificamente, nona e décima *leis da natureza*.

Para Aristóteles, ao contrário, "*pleonexia*" (pode ser traduzido por "*concupiscência*") refere-se à disposição para preferir e escolher, entre ações alternativas, aquela que proporciona, a quem a pratica, o maior prazer imediato. É um tipo de *excesso* ou *desregramento* onde o *desejo do que é agradável* é o mais forte. Cf. *Ética a Nicômaco*, II, 7, 1107 b5 e III, 15, 1119 b1-15.

<sup>112</sup> O termo não é hobbesiano, mas Hobbes faz uso de uma metáfora semelhante: a *capacidade de uma pessoa ver* as boas ou más conseqüências das ações.

<sup>113</sup> A concepção de *probabilidade objetiva* é a de uma probabilidade de natureza *estatística* (genericamente em oposição a outra, de natureza *indutiva*), que diz respeito às leis que regem processos aleatórios quando são avaliadas numericamente as freqüências estáveis, em longo prazo, de ocorrência de eventos; é *objetiva* porque tais freqüências são consideradas como fatos da

peessoas considerarem aquelas conseqüências mais longínquas como insignificantes ou com valor desprezível.<sup>114</sup>

### 3.5. O “egoísmo psicológico”.

Aceitando-se o esquema explicativo das ações apresentado anteriormente, segue-se que sempre é possível atribuir um tipo de motivação egoísta como explicação do comportamento. E assumindo o papel do desejo de autopreservação também como um desejo capaz de, em última análise, motivar todas as ações, isto é, como necessário para a ponderação de todos os fatores relevantes para a decisão (quer dizer, nas ações deliberadas), seja para poder atribuir-lhes peso relativo, seja para poder decidir na ausência de fatores relevantes,<sup>115</sup> não é possível que alguém *tenha interesse em fazer* algo que crê ser em seu próprio detrimento. Isto não exclui a possibilidade de tal pessoa, numa dada situação, *querer fazer* o que crê não ser o melhor a fazer.

Neste sentido, seguindo a psicologia dos agentes a que se aplicam estas considerações, pode-se dizer que “todas as [suas] ações são sempre motivadas pela preocupação por seu melhor interesse próprio ou maior bem”.<sup>116</sup> Este tipo de teoria explicativa é denominado genericamente de “egoísmo psicológico”. Vale notar que, considerando que o egoísta *prefere* - o interesse que o define como um agente deste tipo funciona como um desejo de 2ª ordem (ver nota 26) - não fazer o que crê não ser em seu melhor interesse próprio, mas pode, em certas circunstâncias, não ser forte o suficiente para deixar de desejar fazer o que não crê ser racional fazer, o conceito de *akrasia* (“fraqueza da vontade”) - como Aristóteles a analisou, há também sempre uma crença associada ao

---

natureza independentes do conhecimento que alguém possa ter deles, ou da evidência para apoiá-los. Para uma apresentação da emergência histórica da probabilidade como um conceito dual, ver Ian Hacking, *The emergence of probability*, cap. 2. Uma exposição algo detalhada do contraste entre os conceitos de probabilidade *objetiva* e *subjetiva* pode ser encontrada em Karl Popper, *Realism and the aim of science*, parte II, cap. 1.

<sup>114</sup> Estou pensando mais especificamente na seguinte passagem: “Mais d'autant que les hommes ne peuvent dépouiller entièrement cet appétit brutal, qui leur fait préférer les biens présents (quoi que suivis infailliblement de plusieurs accidens imprévus) aux futurs ...” *De cive*, III, XXXII. Tradução de Samuel Sorbière, revista por Hobbes.

<sup>115</sup> Para a definição do termo *categórico*, quando aplicado a “desejo”; ver 1.10.

<sup>116</sup> Kurt Baier, *Egoism*, in *A companion to ethics*, p. 197.

desejo que o *acrático* tem de fazer algo que sabe, por deliberação, não ser o melhor para escolher fazer -<sup>117</sup> não é incompatível com a tese motivacional genérica do “egoísmo psicológico”.

Aceitar esta compatibilidade supõe aceitar pelo menos dois traços importantes: a tese do “egoísmo psicológico” (a) não expressa, no sentido que é legítimo atribuir a ela, uma generalização sobre os impulsos psicológicos pelos quais os agentes conscientemente são movidos; não se trata então de uma teoria psicológica monista, falsa é claro e (b) não exclui a possibilidade de conflito entre impulsos (as paixões, como exemplo nuclear) para agir, justamente porque um agente pode ser levado por um impulso a fazer o que, nas circunstâncias em que se encontra, não crê ser racional fazer (já sabendo ou ainda não sabendo o que é racional fazer); também não se trata então de uma psicologia moral que torne a aplicação de um princípio ético ou impossível ou inútil.<sup>118</sup>

Explicando uma outra possibilidade: por algum mecanismo psicológico de “auto-engano”, alguém pode continuar a justificar sua ação como motivada por sua preocupação com seu maior bem, i. e., pela preferência que o define como um agente egoísta, ainda que esteja errado em julgar a ação que escolhe como a melhor. Isto é possível porque quem julga que fazer algo é melhor para si está sempre baseado numa *opinião* de que agir de um modo determinado é o melhor que tem a fazer por si mesmo.

Para o caso da psicologia moral hobbesiana, essa opinião pode resultar de um cálculo acertado ou não das conseqüências da ação que se escolhe. No primeiro caso, temos uma opinião ponderada a respeito da adequação racional de fins e meios; no segundo, a ponderação das conseqüências é interrompida ou mesmo distorcida por uma crença sustentada mais fortemente graças à força de alguma paixão. Como veremos no próximo capítulo, é justamente a presença de um elemento cognitivo nas paixões que explica um tipo importante de erro ou insuficiência da razão para orientar corretamente a deliberação.

---

<sup>117</sup> Ver *Ética a Nicômaco*, VII, 5, 1147a30-1147b5.

<sup>118</sup> Estas considerações visam “corrigir” o “egoísmo psicológico” a fim de torná-lo imune pelo menos a algumas das críticas feitas por Sidgwick; ver *The methods of ethics*, livro I, cap. 4, pp. 39-42..

#### 4. TEORIA DA RACIONALIDADE

*"Como a razão é fria e indiferente, não é um motivo da ação, e somente dirige o impulso recebido do apetite ou da inclinação, mostrando-nos os meios de lograr a felicidade e de eludir a miséria"*

David Hume, *An inquiry concerning the principles of morals*

*"As paixões, ... segundo afirma padre Malebranche, justificam-se todas a si mesmas, e parecem razoáveis e proporcionais a seus objetos, na medida em que continuamos a senti-las."*

Adam Smith, *The theory of moral sentiments*

##### 4.1. A razão como cálculo.

Para Hobbes, a *razão* é uma faculdade que funciona unicamente como um *cálculo* (*reckoning*) cujas operações básicas são a *adição* e a *subtração*:

*"Quando alguém raciocina, nada mais faz do que conceber uma soma total, a partir da adição de parcelas, ou conceber um resto a partir da subtração de uma soma por outra; o que (se for feito com palavras) é conceber da consequência dos nomes de todas as partes, para o nome da totalidade, ou dos nomes da totalidade e de uma parte, para o nome da outra parte. (...) Estas operações não são características apenas dos números, mas também de toda a espécie de coisas que podem ser somadas juntas e tiradas umas das outras. (...) Em suma, seja em que matéria for que houver lugar para a adição e para a subtração, há também lugar para a razão, e onde aquelas não tiverem o seu lugar, também a razão nada tem a fazer."<sup>119</sup>*

Portanto, se for possível julgar as ações, desde o ponto de vista da sua racionalidade, como *certas* ou *erradas*, as noções de *adição* e *subtração* devem ser aplicáveis também nesse domínio.

---

<sup>119</sup> *Leviathan*, 5, 18.

#### 4.2. Sobre a deliberação ou raciocínio prático.

Deixando de lado o problema da transformação do "discurso mental" em "discurso verbal" na filosofia da linguagem de Hobbes, podemos dizer que um raciocínio é uma "cadeia de pensamentos" ordenada visando descobrir os meios ("causas") para obter algo ou as conseqüências ("efeitos") produzidas por alguma coisa.<sup>120</sup> No caso dos raciocínios relativos às ações, esta alternativa reduz-se à seguinte:

i) dado um fim, i. e., algo que é desejado, busca-se descobrir o que deve ser feito para alcançá-lo;

ii) dada uma ação, busca-se descobrir quais conseqüências ela trará.

Como resume Hobbes, um raciocínio (prático ou não) é um "método de procura" e descoberta de meios para alcançar um fim ou de fins alcançáveis por um meio. Numa deliberação ou raciocínio prático,<sup>121</sup> ambas as etapas descritas nos itens *i* e *ii* estão envolvidas, simplesmente porque quem delibera, examina as conseqüências de uma ação (ou da sua omissão) não somente para saber se ela é capaz de produzir o que deseja e a que custo, mas também para querer realizá-la ou omitir-se de realizá-la. Deliberações são processos informados de formação de desejos e aversões:

*"Quando surgem alternadamente no espírito humano apetites e aversões, esperanças e medos, relativamente a uma mesma coisa; quando passam sucessivamente pelo pensamento as diversas conseqüências boas ou más de uma ação, ou de evitar uma ação; de modo que às vezes se sente um apetite em relação a ela, e às vezes uma aversão, às vezes a esperança de ser capaz de praticá-la, e às vezes o desespero ou medo de empreendê-la; todo o conjunto de desejos, aversões, esperanças e medos, que vão se desenrolando até que a ação seja praticada, ou considerada impossível, leva o nome de deliberação."<sup>122</sup>*

Relativamente às ações, são as suas conseqüências que fornecem a base para o cálculo realizado numa deliberação. As conseqüências são avaliadas como *boas* ou *más* para quem realizou a ação da qual resultaram. Para que possam ser *somadas* e *subtraídas*, é

<sup>120</sup> Cf. *Leviathan*, 3, 8-9.

<sup>121</sup> Os termos "deliberação" e "raciocínio" não são usados como sinônimos por Hobbes; no entanto, a diferença diz respeito apenas ao tipo de termo lingüístico usado para basear a "cadeia de pensamentos"; cf. *Leviathan*, 6, 29.

<sup>122</sup> *Leviathan*, 6, 28.



necessário primeiro transformá-las em "coisas semelhantes", passíveis de comensuração. Esta condição é satisfeita se é atribuída a cada conseqüência uma grandeza numérica que indica a sua *utilidade*. Além disso, sempre que prevista com base na *experiência*, cada conseqüência de uma *ação particular* ingressa no cálculo como um resultado apenas *provável*:

*"Dado que na deliberação os apetites e aversões são suscitados pela previsão das boas ou más conseqüências e seqüelas da ação sobre a qual se delibera, os bons ou maus efeitos dessa ação dependem da previsão de uma extensa cadeia de conseqüências, cujo fim muito poucas vezes qualquer pessoa é capaz de ver. Mas até o ponto em que se consiga ver que o bem dessas conseqüências é superior ao mal, o conjunto da cadeia é aquilo que os autores chamam bem manifesto ou aparente. Pelo contrário, quando o mal é maior do que o bem, o conjunto chama-se mal manifesto ou aparente. De modo que quem possuir, graças à experiência ou à razão, a maior e mais segura capacidade de prever as conseqüências é quem melhor é capaz de deliberar; e é quem mais é capaz, quando quer, de dar aos outros os melhores conselhos."*<sup>123</sup>

Este trecho é central para entender outro constituinte nuclear do modelo hobbesiano de deliberação e escolha racionais. Assim, pode-se interpretar a racionalidade do agente hobbesiano (leia-se: o modelo de agente pressuposto nas teorias ética e política de Hobbés) como a capacidade de fazer o cálculo da chamada *utilidade esperada*. Este cálculo pressupõe a atribuição de valores (positivo ou negativo) de utilidade a cada um dos prováveis resultados ou conseqüências das ações.

Segundo uma certa concepção, a probabilidade estimada também tem uma medida subjetiva: a probabilidade de cada conseqüência de uma ação é estimada subjetivamente por cada indivíduo.<sup>124</sup> Hobbes está afirmando aproximadamente o mesmo, quando diz que:

<sup>123</sup> *Ibid.*, 6, 29.

<sup>124</sup> Na *teoria da escolha racional* - ou, para ser mais preciso, na parte desta teoria chamada *teoria da decisão* - a chamada "probabilidade subjetiva" é parte do modelo de "escolha sob condições de incerteza". Ver John Harsanyi, *Advances in understanding rational behavior*, in J. Elster (ed.), *Rational choice*, p.87.

"(...) as coisas que estão para vir não têm existência alguma, sendo o futuro apenas uma ficção do espírito (...) E muito embora se denomine prudência quando o acontecimento corresponde a nossa expectativa, contudo, em sua própria natureza, nada mais é do que suposição. Pois a previsão das coisas que estão para vir, que é providência, só compete àquele por cuja vontade as coisas devem acontecer. Dele apenas, e sobrenaturalmente, deriva a profecia."<sup>125</sup>

#### 4.3. Análise do conceito de utilidade.

Aos objetos que proporcionam prazer, também se aplica um outro predicado:

*Útil* se diz de todo e qualquer objeto que alguém julga como bom para garantir-lhe a continuidade da experiência psicológica de sentir prazer; úteis são as coisas que nos agradam no presente e que ao mesmo tempo são meios para fins mais longínquos, coisas que nos agradarão no futuro.<sup>126</sup>

Esta definição do predicado "*útil*" (e do tipo de objeto a que se refere, isto é, a propriedade *utilidade*) pode ser usada para complementar a análise dos juízos de valor apresentada no início do capítulo anterior. Só agora podemos fazê-lo porque fazer tal juízo depende de uma ponderação da *utilidade esperada*, quer dizer, a utilidade de "coisas que estão por vir". Para Hobbes, um juízo "*x é bom*" é sempre um juízo com estrutura lógica intensional, isto é, ele não pode ser reduzido a uma mera atribuição de predicado, mas deve ser analisado como também expressando condensadamente uma *atitude proposicional*: quem afirma que "*x é bom*" está informando algo sobre o prazer que *espera* auferir de *x*; tal informação é o índice da utilidade esperada. Isto ajuda a elucidar mais precisamente o significado da *tese 1*, apresentada em 3.1, sobre a natureza dos juízos de valor.

Coerentemente com uma observação feita por Hobbes no *Leviathan*, pode-se dizer que "*útil*" é apenas um dos tantos "sinais aparentes" por meio dos quais é possível

<sup>125</sup> *Leviathan*, 3, 10. Note-se, por exemplo, a "coincidência" da primeira afirmação de Hobbes nesta passagem com o que diz o economista matemático Kenneth Arrow: "the uncertainty of the consequences ... is understood to be that existing in the mind of the chooser". *Alternative approaches to the theory of choice in risk-taking situations*, *Econometrica*, nº 19, oct. 1951. Citado por Marchetti in *Risco e decisão em investimento produtivo*, p. 75, nota 7. Este artigo de Arrow é um clássico da literatura sobre "decisões sob condições de incerteza" ("*decision under uncertainty*").

<sup>126</sup> Cf. *De cive*, I, X, 85 e *De homine*, 7, 6.

indicar que algo "promete o bem", quer dizer, um dos "nomes" que se dá aos objetos a fim de "testemunhar" que se os julga como bom para (isto é, "capaz de promover") a autopreservação.<sup>127</sup>

<sup>127</sup> Cf *Leviathan*, 6, 24; *De cive*, III, XXI, 119 e Balthazar Barbosa Filho, *op. cit.*, pp. 67-68. Mas o que é um *sinhal aparente*? Se alguém observou que a ocorrência de alguma coisa foi repetidamente sucedida pela ocorrência de outra, se por isso "acostumou-se a ver os mesmos antecedentes seguidos pelos mesmos conseqüentes", sempre que "[ver] ocorrer algo semelhante ao que tinha visto antes, [esperará] (terá a expectativa) que a isso se siga o mesmo que então se seguiu." *De homine*, 4, 7. Também chamará antecedente e conseqüente como *sinais* aparentes (percebidos pelos sentidos) um do outro, da maneira como nuvens carregadas são sinal de chuva no futuro e a chuva é sinal de nuvens carregadas no passado. Mas, diferentemente dos sinais da linguagem, estes não são arbitrários, mas *naturais*. Há uma relação epistemológica entre eles, estabelecida pelo *hábito de esperar* a ocorrência de um a partir da observação do outro. A observação de um passa a ser aceita como um tipo de evidência a favor da ocorrência futura do outro. *Evidência* passa a ser considerada como uma questão de inferir algo a partir de outra coisa: numa imagem, consiste em uma coisa apontar para além de si mesma (do seu conceito).

O conceito de evidência que aparece aqui é aquele que se convencionou chamar de "evidência interna", a partir de seu uso pelos autores da *Lógica* de Port Royal (1662). Hobbes é considerado como um dos verdadeiros precursores do emprego do conceito moderno de *probabilidade* (cujo surgimento histórico pode ser datado: por volta de 1660). Segundo Hacking - ver *Op. cit.*, capítulos 4 e 5 - Hobbes de fato tentou determinar tal conceito, embora não o tenha conseguido com a clareza e fineza conceitual de Pascal, por exemplo. Numa passagem do *De homine* (escrito em 1640 e publicado em 1650), Hacking percebe, para além das simples considerações epistemológicas - ainda escolásticas - de Hobbes sobre os *sinais*, uma imbricação entre *um novo conceito de evidência* (a *evidência indutiva*) e o de *freqüência* de ocorrência (da sucessão entre sinais):

"Os homens pensam vulgarmente que a diferença de sabedoria entre eles resulta da aquisição de sinais a partir da experiência, compreendendo-se comumente por tal aquisição toda a habilidade ou poder cognitivo de um homem. Porém, isto é um erro; pois esses sinais são apenas conjecturais, e conforme tenham falhado com maior ou menor freqüência, a sua garantia será maior ou menor, mas nunca será completa e evidente. Pois embora até hoje se tenha sempre visto o dia seguir-se à noite, não se pode daí concluir que isso ocorrerá ou tenha ocorrido eternamente. A experiência nada permite concluir universalmente. Se os sinais acertam em vinte vezes contra uma, *é nessa proporção que se pode apostar na ocorrência do evento*; mas não se pode concluir isso como uma verdade." *Ibid.*, 4, 10. Grifo meu. Quero acrescentar que a noção hoje técnica de *coeficiente de aposta*, por meio da qual se constrói a concepção normativa de *escolha racional* sob diferentes condições de risco e incerteza, já está aí presente.

#### 4.4. A utilidade é uma medida de expectativa.

Penso que uma análise adequada da definição do predicado "útil" também exige as seguintes considerações. Por um lado, se alguém avalia  $x$  como bom, a base de tal avaliação é o fato de desejar  $x$ . Mas se *desejar* pode ser uma atividade avaliativa, poderia também funcionar como *fonte* única de atribuição de valor? Se isto for assim, então, como primeira consequência, não pode existir nenhum  $x$  que tenha um *valor intrínseco*, isto é, algo que é desejado porque é bom e que não é bom simplesmente porque é desejado. Além disso, dado que uma atividade avaliativa é uma atividade de medição, o que está sendo medido nas avaliações em questão? A intensidade dos desejos, o grau de satisfação (intensidade de prazer) esperado pela realização de tais desejos? Ou ainda, o que se compara quando se avalia a importância relativa de dois objetos  $x$  e  $y$  concluindo, por exemplo, que  $x$  é melhor que  $y$ ?

Na definição de utilidade está enunciado que algo é avaliado como útil se e somente se é julgado como gerador de prazer. Podemos subentender que a extensão do prazer esperado está refletida na medida da utilidade atribuída a um dado objeto. A utilidade de  $x$  é então uma medida da satisfação que se espera obter pela realização do desejo de  $x$ .

Quando alguém julga alguma coisa como *útil*, o que está de fato fazendo é uma *previsão*. Está fazendo, com base na memória que tem de coisas passadas semelhantes, um prognóstico sobre o prazer que espera obter no futuro por meio da fruição de uma coisa determinada. Pode-se dizer que baseou o seu prognóstico em algum *signal* (no caso, uma propriedade natural qualquer) presente que espera ser capaz de produzir prazer no futuro. Mas não há uma relação necessária de sucessão entre sinais naturais.

Há outra operação, também envolvendo risco: a ponderação da expectativa de obtenção da coisa desejada. Esta expectativa está associada ao "custo" do meio utilizado por alguém para conseguir o que quer. Nas condições de extrema insegurança do *estado de natureza*, essa operação passa a ser determinante para a escolha racional de uma ação. Assim, o quanto cada um estará disposto a apostar visando "ganhar" dos outros indivíduos nas diferentes interações dependerá basicamente de dois fatores: a utilidade que atribui ao que quer obter (o que depende, em parte, da expectativa que tem do "efeito" de sua futura fruição, quer dizer, a expectativa associada à sua eficácia na produção de prazer) e a

expectativa que tem de obtê-lo (o que inclui, em parte, a ponderação do risco que corre para obtê-lo).

Estas questões serão consideradas mais detidamente no próximo capítulo. Elas dizem respeito a operações que os homens realizam não apenas como "criaturas meramente sensíveis", isto é, criaturas que "desejam e fazem tudo o que [imediatamente] melhor lhes agrada",<sup>128</sup> mas como seres dotados de razão, isto é, seres capazes de aplicar a si mesmos regras [expressas por enunciados de *deveres*] que governam [ao introduzirem limitações à liberdade de comportamento característica da "condição natural" dos homens] suas escolhas de ação, o que é tornado visível por regularidades comportamentais observáveis.

#### 4.5. Utilidade e expectativas de poder.

Podemos agora combinar a análise anterior dos juízos de valor com esta última versão do significado de proposições valorativas, estabelecendo uma relação entre elas. Quando alguém julga que algo é bom, não pode estar somente querendo dizer que deseja tal coisa. Esta exigência é derivada da necessidade dos juízos de valor poderem ser comparados segundo uma medida que ajuda a estabelecer um critério de escolha entre coisas desejadas. Isto é, os juízos de valor devem poder ser hierarquizados a partir de uma escala de preferências do próprio indivíduo.

Assim, a *utilidade* é o que torna possível a comparação entre preferências, ou ainda, o que torna possível estabelecer, por via indireta, uma escala de desejos segundo a intensidade. "Via indireta" é um modo fróuxo de abreviar algumas relações conceituais. Isto é, segundo esta concepção, quando alguém diz que *prefere x a y* (*x é mais desejado que y*) ou que *x é melhor que y*, está baseando o seu juízo na pressuposição (o juízo pressuposto é uma condição de verdade, ou do significado, do primeiro) de que *x é mais útil que y* (o que indica que *x* lhe proporciona mais prazer que *y*, ou ainda maior satisfação que *y*).

---

<sup>128</sup> *De cive*, prefácio; trad. de Renato Janine Ribeiro.

Se uma coisa pode ou não ser a causa que tem como efeito uma produção de prazer é algo que não depende do juízo ou consideração de alguém. Existe uma relação objetiva de causação, que torna verdadeira ou não a previsão do prazer esperado. A medida da utilidade atribuída a um objeto  $x$  é proporcional à satisfação que se espera obter com a fruição de  $x$ . Ora, a satisfação que se espera obter depende de dois fatores: a necessidade que se julga ter da coisa desejada e o poder que se julga ela ser capaz de prover. Isto porque é do interesse de cada pessoa que a satisfação dos seus desejos presentes seja também um caminho para a satisfação de seus desejos futuros.

Uma coisa desejada, enquanto considerada como um meio ("caminho") para a satisfação de desejos futuros, é o que Hobbes denomina como um *poder*. O termo assinala a capacidade de tal coisa de ser usada para alcançar um outro objeto do desejo. Generalizando, cada coisa é avaliada conforme o que se espera obter pelo uso de seu poder, pelo que pode causar de satisfação imediata e também gerar de satisfação futura, como causa mediata. Se entendemos "poder" neste sentido abrangente, podemos dizer que cada coisa é avaliada pelo que se espera obter do uso de seu poder.

#### 4.6. Sobre a diversidade das concepções individuais de bem.

É um fato a grande variedade dos desejos existentes entre os seres humanos. Parte da explicação para este fato pode ser construída com elementos da própria "psicologia materialista". Os "apetites de coisas particulares" derivam, em última análise, do efeito de cada uma destas coisas no corpo dos homens. Ora, "a constituição do corpo de um homem" se modifica constantemente; logo, é impossível que, a cada momento, uma coisa provoque exatamente o mesmo efeito - isto é, o mesmo tipo de prazer com a mesma intensidade - no corpo de um homem. A conclusão vale igualmente se compararmos os efeitos das coisas particulares distributivamente sobre cada um dos seres humanos.

Porém, outra parte da explicação é a seguinte. As escalas de preferências individuais também podem ser diferentes. Uma escala de preferência resulta da aplicação de algum critério (ou padrão) de comparação entre os diferentes "apetites de coisas particulares" de uma pessoa; se alguém *prefere  $x$  a  $y$*  é porque, segundo algum critério,  $x$  é

79. Segundo algum critério!

Handwritten notes on the left margin:

... a cada momento, uma coisa provoque exatamente o mesmo efeito - isto é, o mesmo tipo de prazer com a mesma intensidade - no corpo de um homem. A conclusão vale igualmente se compararmos os efeitos das coisas particulares distributivamente sobre cada um dos seres humanos.

... a cada momento, uma coisa provoque exatamente o mesmo efeito - isto é, o mesmo tipo de prazer com a mesma intensidade - no corpo de um homem. A conclusão vale igualmente se compararmos os efeitos das coisas particulares distributivamente sobre cada um dos seres humanos.

73  
mais útil que y. Mas de que dependeria o juízo de comparação de utilidades? Depende da função de utilidade que esta pessoa usa para atribuir utilidade a um determinado objeto (coisa ou estado de coisas). Uma função de utilidade associa a cada objeto desejado um valor, uma determinada utilidade. A definição de uma função de utilidade individual depende da concepção de bem de cada indivíduo. Os valores de uma função de utilidade são índices do quanto os objetos desejados são esperados agradar ou desagradar (aqui podemos pensar em valores negativos) alguém. Se uma pessoa avalia o quanto algo lhe agrada ou não com base na extensão do prazer sensual que tal objeto lhe "promete", então sua função de utilidade é diretamente proporcional ao valor da variável "deleite sensual".

Se, em vez disso, por buscar principalmente ser reconhecido por sua glória - independentemente do seu poder e mérito reais -, alguém avalia o quanto de prazer algo lhe promete simplesmente pelas lisonjas que pode esperar dos outros, bem como pelo poder e mérito que poderá atribuir a si mesmo, então a função de utilidade que usa para medir as vantagens destas conseqüências depende diretamente da variável, digamos, "imaginação (infundada) de seu próprio poder, capacidade e mérito"; descrição abreviada por Hobbes com o termo "glorificação", um tipo de alegria (ou prazer do espírito) resultante da expectativa provocada pela previsão das conseqüências de supostas ações heróicas.<sup>129</sup>

Não se pode excluir a existência de funções de utilidade que sejam, na verdade, dependentes de mais de uma variável; nem a existência de funções de utilidade que não podem ser "combinadas", por conterem variáveis incomensuráveis, que não podem ser "pesadas" umas relativamente às outras. Teremos, assim, também uma diversidade interna à concepção individual de bem; o bem individual pode ser metaforicamente concebido como um cesto preenchido por uma pluralidade de itens (os diferentes tipos de prazer): se tais itens podem ser representados na forma de índice (representado por uma única função), a pluralidade é constitutiva; caso contrário, ela é competitiva.<sup>130</sup>

<sup>129</sup> Ver *Leviathan*, 6, 25-27 e *De homine*, 9, 1.

<sup>130</sup> Estes conceitos são tomados de empréstimo da análise de Amartya Sen, *The standard of living: lecture I, concepts and critiques* in G. Hawthorn (ed.), *The standard of living*, pp. 1-19. Podemos também conceber um exemplo onde diferentes tipos de prazer, embora não incomensuráveis em princípio, determinam funções de utilidade suficientemente diferentes (ainda que possam ser, sob alguma descrição, semelhantes) para tornar, na prática, muito difícil a comparação dos juízos de valor feitos com base nelas. Este é precisamente o exemplo concebido por Wittgenstein (tomo-lhe emprestado apenas o exemplo, que aparece originalmente como parte [da análise de Wittgenstein] dos juízos estéticos):  
de sua análise

#### 4.7. Juízos de valor subjetivos como “razões para agir”.

Podemos agora sublinhar alguns pontos. Dizer que "um agente crê que  $x$  é bom" ou que "ele julga  $x$  como bom" equivale a dizer que ele "deseja  $x$ ". E, sempre que agir guiado por tal desejo, pode oferecer, como *razão para agir* para obter  $x$ , o simples fato de desejar  $x$ . Uma tal razão para agir, baseada em juízos de valor subjetivos, é classificável como *uma razão relativa ao agente (agent-relative reason)*.

Em oposição, poderíamos falar de *uma razão neutra em relação ao agente (agent-neutral reason)*, conceito que pode ser analisado do seguinte modo: se o juízo de valor, feito por  $A$ , de que " $x$  é bom" implicar não somente uma razão para  $A$  agir, mas também uma razão para *toda pessoa* (a quem isto esteja ao alcance) agir para preservar, promover, proteger ou assegurar  $x$  para  $A$ , este juízo deixa de ser subjetivo e fornece uma razão para agir neutra em relação a todos estes agentes, incluindo o próprio  $A$ .<sup>131</sup>

Considerada até aqui, e classificada desde um ponto de vista meta-ético, a teoria ética de Hobbes pode ser classificada como uma *teoria hedonista subjetivista*. *Hedonista* porque afirma que bom é o que produz prazer. *Subjetivista* porque a medida do prazer - e, conseqüentemente, da bondade das coisas - é sempre relativa a um indivíduo, não se podendo atribuir a nenhum objeto um valor *intrínseco* (ver 4.4). Não se exclui, no entanto, uma suposição adicional: a possibilidade de um acordo de diferentes indivíduos sobre a utilidade de um determinado objeto. Ela é, isto sim, incompatível com qualquer concepção forte de *razão neutra em relação ao agente*.<sup>132</sup>

---

"Suponham que encontramos alguém na rua e essa pessoa nos diz que perdeu o seu maior amigo, com uma voz que exprime de um modo extremo a sua emoção ... Suponham então que perguntamos: 'Que semelhança existe entre admirar essa pessoa e gostar de sorvete de baunilha?' Comparar as duas coisas parece quase repugnante. (Mas podemos ligá-las através de casos intermediários.) Suponham que alguém disse: 'Mas este é um tipo de prazer bastante diferente'... Existe uma ligação entre ambos os prazeres. Apesar de no primeiro caso a emoção do prazer adiantar pouco para o nosso juízo". *Wittgenstein's lectures and conversations*; pp. 32-33 na tradução portuguesa.

<sup>131</sup> Cf. Gaus, *Value and justification*, p. 191.

<sup>132</sup> Quer dizer, ela exclui a interpretação "altruísta" de "valor objetivo": objetividade, entendida corretamente, significa *imparcialidade*, e só pode haver imparcialidade se é dado um igual peso ao ponto de vista dos outros, quer dizer, sob condições apropriadas, os *meus* juízos de valor, pelo simples fato de serem *meus*, não fornecem razões para agir superiores às fornecidas pelos juízos dos outros. Ao implicar num tipo extremo de dissociação de um sujeito dos *seus próprios* valores, pelo menos como pressupõe as teorias que aceitam esta concepção de objetividade, tal interpretação pode ser dita pressupor uma "concepção *impessoal* de bem". Ver Gaus, *op. cit.*, pp.192-193.



O *bem*, se interpretado como a continuidade *individual* do prazer, admite uma pluralidade de *medidas privadas*. Mas o que se diria de algo que consistisse numa ameaça permanente à continuidade do prazer de qualquer indivíduo? Segundo Hobbes, o que cada ser humano mais quer é “evitar a destruição de sua natureza pela violência”, e isto é o que explica por que cada um concebe como racional criar condições que possam, durante todo o tempo de sua vida, consistir-se em garantia contra tal destruição.<sup>133</sup> Estas condições estão enunciadas em preceitos da razão, que não são suficientes para garantir a abediência dos homens. Este ponto será desenvolvido na seção 4.16.

#### 4.8. A concepção hobbesiana de razão prática.

Há uma especificidade na concepção hobbesiana de *razão prática* com relação à concepção padrão da moderna teoria econômica. Nesta, a exigência de maximização deixa sem especificar quais são as características da medida a ser maximizada, isto é, não se determina quais objetos devem contar como *objetos de preferência*. Apenas se supõe a existência de uma medida precisa das preferências: a utilidade (que é, às vezes, referida mais vagamente pelos termos “vantagem”, “benefício” ou “satisfação”).<sup>134</sup> Pode-se atribuir a Hobbes uma concepção de racionalidade como maximização. Porém, para ele, o valor a ser maximizado é o “benefício próprio” (ou, um valor de segunda ordem: “a conservação própria”), como mostram as seguintes afirmações:

*“(...) jamais se poderá dizer que [são] contra a razão [as ações que] tendem para [o] benefício próprio” e “as ações mais razoáveis são as que melhor conduzem [ao benefício próprio]”*<sup>135</sup>

O que Hobbes está fazendo é acrescentar à idéia de “fazer escolhas procurando obter o melhor que é possível obter” uma especificação de que as preferências pelas diferentes conseqüências sejam sempre baseadas numa “medida auto-direcionada” (expressão de Gauthier). A partir desta exigência adicional, o agente racional hobbesiano

<sup>133</sup> Cf. *Leviathan* 15, 79 e 17, 86.

<sup>134</sup> Cf. Gauthier, *Morals by agreement*, p. 22 e *Moral Dealing*, p. 234, nota 4.

<sup>135</sup> Leia-se: “é racional fazer aquilo que maximiza o benefício próprio”. *Leviathan*, 15, 72.

faz as suas escolhas buscando obter *para si mesmo* o melhor que é possível obter. Isto o define como um (agente) egoísta.

Seguindo Sidgwick, podemos definir a "base sobre a qual Hobbes tentou construir a moralidade" como o princípio do "egoísmo racional" (também denominado como "axioma da prudência racional"): é racional (e correto) para uma pessoa, em quaisquer dadas circunstâncias, agir de modo a produzir a maior felicidade (ou prazer) próprio possível.<sup>136</sup> E, como observa Sidgwick, na teoria ética de Hobbes, o "fim supremo ou standard da conduta correta" do egoísmo ("a maior felicidade própria") não é interpretado como ele próprio o faz, isto é, como "o maior grau de satisfação dos seus desejos" [i.e., como "a maior quantidade de prazer pessoal"], mas como "a autopreservação". Todavia, penso que as diferenças aqui podem ser consideradas, sem maiores implicações, como apenas nominais.

#### 4.9. A busca egoísta da felicidade.

Atentando para a origem do desacordo quanto aos valores entre os homens, constata-se que ele é paradoxalmente gerado pelo mesmo desejo fundamental que, *qua* seres humanos, todos compartilham: o desejo de autopreservação. Contudo, isso implicaria em dizer que existe algo (a autopreservação) que é avaliado como bom ou desejável por todo ser humano. Uma avaliação da forma lógica desta afirmação mostra que se está tentando atribuir ao egoísmo uma doutrina que não só não lhe pertence, como também, se defendida como parte dele, serviria para refutá-lo.

Vimos que a autopreservação é a interpretação hobbesiana do conceito de felicidade. Vimos também que o conceito avaliativo fundamental do egoísmo é sempre expresso por uma relação: "x é bom para y". Igualmente, o princípio fundamental do egoísmo é: "a felicidade própria é o fim último das ações de cada indivíduo". Adequadamente analisado, este princípio tem a seguinte forma lógica (as três expressões abaixo são equivalentes):

---

<sup>136</sup> Cf. *The methods of ethics*, pp. 89 e 381.

- (1) para todo  $y$ , a felicidade de  $y$  é o fim último de cada ação de  $y$ ; ou
- (2) para todo  $y$ , a preservação de  $y$  é boa *para*  $y$  (i. e., boa *relativamente a*  $y$ ); ou ainda
- (3) para todo  $y$ ,  $y$  deve buscar a felicidade de  $y$ .

É um traço definidor da concepção egoísta que a pessoa que busca a felicidade seja a mesma pessoa cuja felicidade é buscada. Se este traço não fosse relevante, seria possível sustentar (considerando uma pessoa  $A$ ) que a felicidade de  $A$  é *intrinsecamente* desejável ou boa, querendo com isto dizer, ela é *objetivamente* boa. Neste caso estaríamos tratando os termos *bom*, *correto* e *deve* como parte de predicados monádicos (não-relacionais). Poderíamos então quantificar universalmente um deles (a mesma quantificação é válida para os demais):

para todo  $y$ ,  $y$  deve buscar a felicidade de  $A$ .

A expressão "autopreservação" significa "a preservação de algum indivíduo em particular"; se determinamos um indivíduo, ela pode passar a significar "a conservação de  $A$ " ou, equivalentemente, "a felicidade de  $A$ ". Assim, teríamos uma interpretação para a idéia mais vaga de que a autopreservação (isto é, a preservação de  $A$ , ou de  $B$ , ou de  $C$ , e assim também para todo e qualquer indivíduo) é o valor fundamental de todos os seres humanos. Mas para chegar a esta conclusão, tem-se que pressupor os conceitos valorativos indicados como predicados monádicos. Ora, isto é algo que a concepção egoísta não pressupõe. [Portanto, uma tal interpretação simplesmente erraria o alvo.]  $\Delta$ lt.

#### 4.10. Os papéis da razão e das paixões na deliberação.

Desejar ou não fazer uma ação é o que Hobbes chama respectivamente de "último apetite ou aversão" da *deliberação*. Uma vez chegado ao fim, o processo de deliberação também põe fim à liberdade que, antes de iniciar tal processo, "tínhamos de praticar ou evitar a ação, conformemente a nosso apetite ou aversão".<sup>137</sup> É no processo de deliberação que "passam sucessivamente pelo pensamento as diversas conseqüências boas ou más de uma ação"<sup>138</sup> ou da sua omissão. E é precisamente o cálculo da *utilidade esperada total* de cada uma destas alternativas que permite suscitar, no fim da deliberação, o apetite ou a aversão informados da ação.

Este cálculo funciona do seguinte modo. Primeiramente, são definidas uma utilidade e uma probabilidade associada a cada conseqüência de uma ação e da omissão desta ação. Representando uma ação genérica por  $x$  e a sua omissão por  $\sim x$ , o cálculo da utilidade esperada - para orientar a escolha - é guiado pela seguinte fórmula:

$$UE(x) = \sum_{i=1}^n p_{ci} \cdot U_{ci}$$

Leia-se: a utilidade esperada da ação  $x$  é igual ao somatório das utilidades esperadas de  $n$  conseqüências de  $x$  ( $c_1, c_2, c_3, \dots, c_n$ ). Para  $UE(\sim x)$  também vale a mesma fórmula.

Se o agente é racional, ele escolhe - pelo menos de acordo com a abordagem *bayesiana* - a alternativa que *maximiza* a utilidade esperada. Mas o resultado pode não depender exclusivamente da escolha de um indivíduo; isto é precisamente o que acontece no estado de natureza, onde os indivíduos necessariamente interagem uns com os outros. Neste caso, dizemos que um *agente estrategicamente racional* leva em consideração o fato de que o ambiente é "povoado" por outros atores, que ele próprio é parte da situação de escolha deles, que sabem do seu comportamento e que o levam em conta nas suas expectativas das conseqüências de alternativas possíveis de suas ações. Como bem assinala Jon Elster, a consideração do comportamento estratégico permite sublinhar o "fato de que os atores humanos não somente fazem as suas escolhas com base nas suas expectativas

<sup>137</sup> *Leviathan*, 6, 28.

<sup>138</sup> *Idem*.

sobre o futuro, ... mas também *com base nas suas expectativas sobre as expectativas dos outros*".<sup>139</sup>

A razão funciona apenas como instrumento de realização do que é ditado pelas paixões. A especialização dos papéis das paixões e da razão nos raciocínios práticos pode ser entendida por meio da descrição do "comportamento lógico" da linguagem que expressa cada um deles:

*"A deliberação se exprime pelo Subjuntivo, que é o modo próprio para significar suposições e suas conseqüências, como por exemplo em Se isto for feito, esta será a conseqüência. (..) A linguagem do desejo e da aversão é Imperativa, como por exemplo em Faze isto ou Evita aquilo."*<sup>140</sup>

"Deliberação" significa, nesta passagem, apenas o processo cognitivo de descoberta das conseqüências de uma ação.<sup>141</sup> Os resultados deste processo são sentenças *condicionais*. Uma ação é escolhida como um meio para alcançar um fim, ao qual não faz sentido aplicar o predicado *racional*, porque a razão não pode ditar à vontade o que esta deve desejar, mas somente o que deve ser feito para alcançar o que é desejado. A operação da razão limita-se ao cálculo das conseqüências, avaliadas a partir do prazer ou desprazer que "prometem". Este traço da racionalidade foi aceito mesmo por Adam Smith, para quem a teoria ética hobbesiana é construída a partir de um grande engano, a saber, a concepção do auto-interesse (o termo de Smith é "*self-love*") como base do valor moral das ações:

*"Mas a razão não pode tornar nenhum objeto particular por si mesmo agradável ou desagradável para a mente. A razão pode mostrar que este objeto é o meio de obter algum outro que é claramente prazeroso ou não, e deste modo torná-lo agradável ou não porque desejado como meio para outra coisa"*<sup>142</sup>

A cadeia de pensamentos para determinar o melhor meio para a satisfação do desejo inicial é desencadeada por este próprio desejo, o qual se transmitirá acompanhando as sucessivas descobertas do que é não somente mais eficaz na produção do fim desejado, como simultaneamente é também aquilo que o agente tem o poder de fazer:

<sup>139</sup> *Ulisses and the sirens*, p. 19.

<sup>140</sup> *Leviathan*, 6, 29.

<sup>141</sup> É importante observar que Hobbes não usa portanto esse termo num único sentido.

<sup>142</sup> *The theory of moral sentiments*, parte VII, seção III, cap. II, p. 320.

"(...) a impressão feita por aquelas coisas que desejamos, ou receamos, é forte e permanente, ou (quando cessa por alguns momentos) de rápido retorno. (...) Do desejo surge o pensamento de algum meio que vimos produzir algo de semelhante àquilo que almejamos; e do pensamento disso, o pensamento de meios para aquele meio; e assim sucessivamente, até chegarmos a algum início dentro de nosso próprio poder. E porque o fim, pela grandeza da impressão, vem muitas vezes ao espírito, no caso de nossos pensamentos começarem a divagar, eles são rapidamente trazidos de novo ao caminho certo."<sup>143</sup>

#### 4.11. As paixões e as falhas na escolha racional.

As paixões (principalmente, o maior ou menor desejo de poder, de riqueza, de saber e de honra) determinam nossas *capacidades cognitivas*, com exceção da *razão* (vista aqui como a capacidade de usar corretamente a linguagem, único *talento adquirido*, isto é, alcançado através de "método e instrução"). As capacidades cognitivas são basicamente a *imaginação* e o *juízo*.

As paixões determinam a intensidade dos desejos e as preferências. Quando fracas, produzem desejos débeis e preferências indiferentes. Determinam também o alcance da imaginação, quer dizer, das cadeias de pensamentos guiadas (com maior ou menor "grandeza da impressão" do fim desejado) pela busca (ou *capacidade de invenção*) "das causas de algum efeito presente ou passado, ou dos efeitos de alguma causa passada ou presente".<sup>144</sup>

"Porque os pensamentos são para os desejos como batedores ou espias, que vão ao exterior procurar o caminho para as coisas desejadas; e é daí que provém toda firmeza do movimento do espírito, assim como toda rapidez do mesmo".<sup>145</sup>

Há um tipo de situação em que a razão pode ajudar a eliminar incoerências dos desejos. Um exemplo simples: alguém que deseja simultaneamente comer um bolo

<sup>143</sup> *Leviathan*, 3, 9. A concepção de deliberação aqui exposta é praticamente a mesma de Aristóteles; isso pode ser confirmado na famosa passagem da *Ética a Nicômaco*, 1112 b 11-28, onde este filósofo compara a deliberação ao procedimento de análise matemática.

<sup>144</sup> *Leviathan*, 3, 10.

<sup>145</sup> *Ibid.*, 8, 35.

imediatamente e guardá-lo para mais tarde, sustenta desejos incoerentes porque *não existe uma ação que seja meio para realizá-los ao mesmo tempo*. Este tipo de incoerência nem sempre é tão direta, como mostra o exemplo de alguém fazendo as seguintes afirmações: eu gostaria de não gostar de comer bolo de creme, mas gostaria de comer bolo de creme porque adoro bolo de creme. Eu gostaria de não gostar disso porque, sendo uma pessoa um tanto vaidosa, creio que é mais importante manter-me esbelto; mas gostaria de ser menos vaidoso.<sup>146</sup>

Este caso pode ser visto como um exemplo de incoerência entre, de um lado, desejos orientados para a gratificação imediata e, de outro, desejos orientados para a gratificação futura. Este tipo de incoerência ajudará a descrever uma causa básica da falha dos agentes que se encontram na situação descrita por Hobbes como o "estado de natureza" em fazer escolhas racionais.

Considerada a possibilidade destas falhas, segue-se que os agentes escolhem, algumas vezes, ações cujas conseqüências, tomadas como um todo, lhes são danosas. Nos casos em que os agentes não são capazes de prever as conseqüências menos próximas das suas ações, o tipo de erro em questão pode ser genericamente descrito como "tomar um bem aparente por um bem real". Um bem aparente pode ser definido como algo que alguém julga ser bom para si, algo que estima lhe ser benéfico com base na informação de que dispõe nas circunstâncias em que faz sua escolha. Mas, além do fato desta informação ser *por princípio* sempre limitada (o agente deve em algum momento decidir interromper a coleta de evidências), a racionalidade do agente também pode ser limitada, quer dizer, na deliberação, elementos em parte apenas não-cognitivos (as paixões)<sup>147</sup> podem de algum modo "deformar" os elementos cognitivos (as crenças e as premissas que as sustentam, suas evidências).

Neste último caso, a razão é limitada porque não é suficiente para a correção de erros. São casos em que alguém demonstra, por alguma coisa, "paixões mais fortes e veementes do que geralmente se verifica nos outros", quando em circunstâncias semelhantes. Esta resposta emocional exagerada a uma dada situação é o que Hobbes denomina genericamente de *loucura*. O importante, para a análise presente, é saber de que tipo de *erro* estamos falando.

<sup>146</sup> O exemplo é de Jon Elster em *Nuts and bolts for the social sciences*, p. 37.

<sup>147</sup> Ver 2.10.

24. [Minha análise partirá de uma discordância. Jean Hampton, ao montar uma interpretação (citada a seguir) que crê ser hobbesiana sobre o fenômeno das “vítimas de um erro biológico”], faz pelo menos duas pressuposições:

“Seus movimentos corporais estão de tal modo perturbados que elas se tornam completamente incapazes de reconhecer o que aumentará seus movimentos vitais (i. e., o que lhes dá prazer), e é em virtude desta massiva falha biológica que elas são ditas irracionais... [Com] este uso da palavra ‘irracional’... Hobbes está dizendo que pessoas doentes estão cometendo um erro relativo a como buscar prazer, ainda que não cometam um erro cognitivo na busca de um objeto do desejo, mas um ‘erro’ biológico absolutamente não-cognitivo no processo de formação de um desejo... Assim, nesta concepção de racionalidade, a razão deve ser considerada como instrumento somente de desejos reais, não de desejos espúrios”.<sup>148</sup>

Não estou interessado em avaliar o ponto de vista desta intérprete sobre o chamado “erro biológico”. As razões para considerar estas pressuposições como inaceitáveis são: (a) um erro totalmente não-cognitivo (talvez se possa dizer isto, com sentido, de um processo corporal *patológico*, comparando-o ao *normal* ou *saudável*) não pode, por si só, autorizar-nos a falar de *um erro sobre* o meio mais adequado (“a ação correta”, diz Hampton, na p. 39) de buscar prazer, e (b) não é compatível com a concepção hobbesiana de racionalidade dizer que a razão tem de ser considerada como instrumento somente de desejos *reais* e não de *espúrios*.

Creio que é possível não só compatibilizar o que sustento com as teorias da motivação, do valor e da racionalidade hobbesianas, como também tramá-las de modo mais simplificado. Para isso, primeiro é preciso aceitar a análise das paixões que as divide entre componentes cognitivos e não-cognitivos. Depois, aceitar que paixões podem funcionar como desejos de 2<sup>a</sup> ordem, estabelecendo preferências e respectivas hierarquias, tornando possível uma análise ao mesmo tempo mais nítida e refinada da deliberação e escolha.

Tomemos o exemplo de duas importantes paixões que predispõem os homens a se comportar como “predadores” ou “competidores” (ver nota 30): a *vanglória* e a *cobiça* (ver 3.4). Paixões *em excesso* (de intensidade ou de duração) são o que Hobbes denomina como “*espécies de loucura*”. Loucuras se manifestam por “comportamentos estranhos e invulgares”, ou ainda, “extravagantes”. Vou deixar de lado o problema de saber se apenas

<sup>148</sup> *Op. cit.*, pp. 40-41. Hampton começa a analisar este ponto na p. 36.



em excesso, ou se também em graus considerados como moderados, a vanglória e a cobiça, quando *dominantes* do comportamento, são geradoras de discórdia. Vou me fixar numa avaliação qualitativa da natureza do erro cuja correção não está ao alcance da razão.

Começemos pelas definições:

(1) *A* é motivado por *vanglória* se e somente se:

(i) *A* tem *esperança* de obter *x* (deseja e crê ser possível obter *x*); e

(ii) *A* sente *alegria* (um "prazer do espírito") ao (iii) se *imaginar* poderoso para obter *x*; e

(por análise de iii) *A* apenas *inventa* ou *supõe* ter suficiente *capacidade* ou *poder* para obter *x* (o que, por definição, implica: se *A supõe*, *A* não sabe que possui tal capacidade ou poder; e se *A inventa*, *A* sabe que não possui tal capacidade ou poder).

(2) *A* é motivado por *cobiça* se e somente se:

(i) *A* tem *esperança* de obter *x*; e

(ii) *A* julga que a uma ação *competitiva* (ver nota 30)<sup>149</sup> é o melhor meio para obter *x*, porque (iii) *A* quer agir de modo a ter, no presente, sempre mais das coisas que lhe dão poder (ver 3.4).

Vejamos como é possível que uma paixão seja fonte de um erro relativo a um meio para buscar prazer, ainda que não seja a base de um engano quanto ao meio para obter um dado *x*, um *objeto do desejo*. O que pretendo sublinhar, conforme as discordâncias recém apontadas, é que não seria correto, pelo menos de modo coerente com a teoria hobbesiana, afirmar que peçoas que buscam *x* movidas por emoções exageradas agem irracionalmente porque não são capazes de reconhecer *x* como um meio de aumentar seu poder.

Seu erro consiste em não *julgar y como* o meio racional de buscar prazer; ele depende, em última instância, da ausência de um desejo de 2ª ordem, a saber, o desejo de adiar a satisfação do desejo de *x* pela razão de julgar *x* apenas como um meio provisório, incerto e, por isso, pouco eficaz de aumentar seu poder e prolongar sua vida. A *prudência* é este desejo de 2ª ordem.

<sup>149</sup> Sugiro que, entre as ações competitivas, seja incluída a que Hobbes denomina como *antecipação*. Ver Leviathan, 13.

Prudência e erro de cálculo

(ii) julgar como meio provisório  
há diferenças entre (a) risco-veloz e (b) julgar? Ou a diferença é sobre simples? tenta um erro de cálculo? Talvez por falta de certeza a prudência = a liberdade.

Qual a diferença, se alguma entre: (ii) não julgar como meio provisório

Combinando a análise acima com a de *motivo*, vista em 1.3, temos:

*A* faz *y* motivado por *cobiça* se e somente se:

- (i) *A* deseja obter *x*;
- (ii) *A* crê que o melhor meio para obter *x* é fazer *y* (uma ação *competitiva* qualquer).

Desta forma, fazer *y* motivado por *cobiça* é irracional porque supõe o erro de não julgar a ineficácia de *y* como meio para garantir poder por mais tempo. Não deixar de fazer *y* é um “erro relativo a como buscar prazer” somente se pressupomos que esta busca deve ser a mais compatível com a garantia de obtenção de poder por mais tempo. Os imprudentes não “respondem” a “como buscar prazer?” com “por mais tempo”, e esta limitação de preferência não pode ser “corrigida” pela razão.

Para o caso da *vanglória*, tal combinação é a seguinte:

*A* faz *y* motivado por *vanglória* se e somente se:

- (i) *A* deseja obter *x*;
- (ii) *A* crê que o melhor meio para obter *x* é fazer *y* (uma ação que lhe dá o prazer de se imaginar poderoso).

Neste caso, o erro também consiste genericamente em buscar prazer escolhendo uma ação que não é a melhor se é visada a satisfação do desejo (de 2<sup>a</sup> ordem) de fazer o que de fato serve de meio para um aumento de poder. Quer dizer, não é prudente supor, muito menos inventar, que se dispõe de poder; fazê-lo consiste no erro - em certos casos, com conseqüências bastante danosas - de tomar a aparência pela realidade, ou melhor, “tomar um *bem aparente* por um *bem real*”.

#### 4.12. "Bem real" e escolha racional.

Um *bem real* é algo que um agente racional continuaria desejando fazer mesmo após corrigir as limitações (cognitivas e não-cognitivas) das condições que determinam a sua escolha. Um bem real para alguém pode ser definido então como uma ação que tal pessoa *quer fazer, e que julga, com boas razões, ser bom para ela fazer*, após ponderar os desejos que considera relevantes para a sua escolha. Estes são desejos de objetos particulares (cuja fruição lhe será prazerosa), (cuja ponderação depende da operação de desejos 2ª ordem).

Uma *ação boa* é, para Hobbes, uma ação que contribui para a realização da paz. Dizer que uma ação representa um bem real para *mim* significa dizer que, nas circunstâncias em que é realizada, é a ação racional, a ação que mais eficazmente produz a paz, quer dizer, que não só *me* garante a fruição de meu atual objeto do desejo, como também a que *me* dá mais poder para realizar futuros desejos. Supõe-se que, se *minha* ação não contribui para a manutenção da paz, se a ação que faço é um fator de geração de discórdia, desconfiança, violência, então haverá *para mim* uma ameaça constante de perder poder, de perder meios que disponho para continuar satisfazendo desejos que possa vir a ter.

É possível escolher fazer *x* - porque se crê que suas conseqüências, como um todo, são boas - e não se conseguir as conseqüências que se deseja ao fazer *x*, ou consegui-las, mas com um balanço negativo de utilidade. Neste caso, há duas explicações para incorrer numa escolha irracional. É possível que alguém creia, com boas razões, que as conseqüências do que escolhe fazer contribuem, consideradas como um todo, para a manutenção da paz, com a qual deixa de ser racional temer a perda de poder pessoal. Vale aqui, portanto, o comentário: "não apenas errar é humano; pode mesmo ser racional fazê-lo, se acontecer de todos os indícios apontarem na direção errada".<sup>150</sup> Se, no entanto, os indícios apontam na direção errada porque isto é justamente *o que se quer* (é o que se chama de *auto-ilusão*), o erro pode ser humano (a natureza humana nos ajuda a precisar do que se trata), mas não se pode tratar de uma *escolha racional*.

<sup>150</sup> Elster, *Nuts and bolts for the social sciences*, p. 41 na ed. brasileira.

cálculo das consequências e = a  
desejo de 2 e ordem?

Prudência?  
é o nome de um  
virtude? Para  
há bens  
e  
maus  
consequências

Esta última alternativa demanda uma explicação mais cuidadosa. Uma ação é julgada boa ou não com base no cálculo de suas consequências. Este cálculo depende da atribuição de utilidade às consequências, valor que é relativo aos *apetites e aversões* suscitados pela previsão das consequências (ver 4.2). Se entendemos que a *prudência* é uma disposição a ponderar desejos (apetites e aversões) segundo uma regra de preferência específica. Tal como já assinalava Platão no diálogo *Protágoras*, um homem que sabe realizar adequadamente as ponderações envolvidas numa deliberação, o homem *prudente*, "sabe pesar" as consequências de ações alternativas porque não se detém apenas na proporção entre coisas agradáveis e desagradáveis:

"(...) um homem que sabe pesar põe num prato da balança as coisas agradáveis e no outro as coisas desagradáveis, dum lado as coisas que estão próximas, do outro as coisas que estão afastadas, e vê para que lado pende a balança"<sup>151</sup>

Desta forma, a vantagem que cada um espera obter do que pretende fazer, acrescenta Sócrates no diálogo, é determinada por uma operação que necessariamente inclui uma capacidade para *medir corretamente* prazeres e dores. "Medir corretamente" quer dizer: fazer juízos verdadeiros a respeito da medida dos prazeres e dores associados às consequências das ações que realizamos. Fazer tais juízos pressupõe a supressão de fontes de erro: a principal destas sendo justamente a de se deixar enganar pela *aparência* das coisas. Assim como os mesmos objetos nos aparecem como maiores de perto e, menores, quando vistos de longe, também pesar inadequadamente as consequências de ações - medindo-as em excesso, quando mais próximas, e, em falta, quando mais afastadas - é uma fonte de erro.

Se alguém objetasse dizendo que "o prazer atual difere muito do prazer ou dor futuros", Sócrates imediatamente replicaria que não podem diferir *enquanto* prazer e dor, mas senão "por uma coisa diferente do prazer e da dor"<sup>152</sup> (hoje diríamos: se diferem, diferem "sob uma outra descrição"), a saber, como os avaliamos por sua maior ou menor proximidade temporal. A medida da vantagem que um homem prudente espera obter do que pretende fazer depende da aplicação de uma regra. É graças à regra que ele escolhe ações que são bens reais, o que fica mais evidente quando resiste ao medo que lhe inspiram

<sup>151</sup> P. 76 na ed. portuguesa.  
<sup>152</sup> Op. cit., pp. 75-76.

"aplica regra" é confundido  
com a doutrina da afecção  
e da psicologia determinista  
de Hobbes.

Em Hobbes se ficou de cima a 2ª alternativa:  
"Va como um cálculo usado para se reduzir o desejo de não sofrer discórdia" e o desejo de segurança, em part, meu risco.

males próximos, porque reconhece que lhes atribuir um peso excessivo por sua proximidade é um passo para aumentar os males produzidos pela discórdia. Imagine-se alguém que, temendo ser trapaceado num contrato, gostaria de trapacear primeiro; mas que resiste a trapacear agora, porque isto aumentará suas chances de ser futuramente trapaceado. Diferentemente, poderia resistir ao medo de ser trapaceado agora e deixar de trapacear primeiro porque não quer contribuir para aumentar a discórdia, mesmo que um tal aumento não implicasse necessariamente no risco maior de ser futuramente trapaceado?

Hobbes não discorda de afirmações sobre a natureza das virtudes. Por exemplo: se for suficiente dizer que um homem prudente é, *ipso facto*, virtuoso ("x é virtuoso" se e somente se "x possui uma virtude no sentido de 'uma disposição interior da alma'" é uma definição aceita por Hobbes), isto é, se a prudência é tomada como uma virtude, então é a última possibilidade, referida no parágrafo anterior, que apresenta a razão pela qual o homem prudente deixa de trapacear, e age virtuosamente. Teoricamente, contudo, Hobbes parece considerar mais realista a suposição de que se deva contar com o *medo* como a principal motivação para as pessoas fazerem o que é correto fazer,<sup>153</sup> o que neste caso quer dizer apenas *agir em conformidade com as leis da natureza*, cumpri-las por imposição daquele a quem autorizaram o mando sobre si mesmas.

No modelo hobbesiano, a felicidade de cada pessoa, a conservação da sua vida, depende das ações que escolhe: fazer o que é realmente bom é o que melhor pode fazer por si mesma. O que não quer dizer que os que fazem as escolhas racionais as façam pelas mesmas razões (isto será precisado por meio do exemplo da *justiça*, na seção 5.13). Uma virtude é predicada de uma ação apenas indiretamente: uma ação é virtuosa se a vontade de

<sup>153</sup> Há um importante aspecto envolvido nesta afirmação. Na escolha teórica feita por Hobbes, não é necessário pressupor uma generalização (psicológica e sociológica) pessimista a respeito de todas as sociedades e pessoas. Dado o valor da paz para "escapar" do estado de natureza, parece razoável atribuir uma atitude conservadora ou cautelosa para guiar as pessoas que querem decidir qual poder instituir para garantir a cooperação social. Tal atitude poderia ser modelada teoricamente pelo uso do *princípio maximin*, que prescreve escolher, supondo a existência futura das piores circunstâncias, a alternativa que traz os melhores resultados: no caso, se supomos um grande número de pessoas comportando-se de modo egoísta em sociedade, o melhor que há para fazer é escolher um tipo de poder capaz de impor, mesmo a pessoas assim motivadas, o cumprimento das leis. Ver Nozick, *Anarchy, state and utopia*, p. 19 na trad. brasileira; e Rawls, *A theory of justice*, pp. 152-154 e 240.

Após  
esta  
definição  
de virtude

Man  
esta  
questão  
de  
indiretamente?

Medo

quem a faz é determinada por alguma virtude. As virtudes cardiais são, para Hobbes, as mesmas da ética platônica: a *prudência*, a *temperança*, a *coragem* e a *justiça*.<sup>154</sup>

Num sentido, ser prudente pressupõe a disposição a não se deixar vencer pelo prazer. Não ser vencido pelo prazer deve ser corretamente entendido como não deixar de fazer o que é melhor para si, ao se admitir deliberadamente que uma conseqüência agradável pode ser boa "sob um aspecto", e ainda assim ser má, sob algum outro aspecto: porque este aspecto depende da variação do intervalo de tempo considerado. Platão habilmente decifra o enigma para nós:

"— Portanto — continuei — se eles disseram: 'Que entendeis, pois, pelo que chamastes ser vencido pelo prazer?' Eis como eu lhes responderia: 'Escutai, nós, eu e Protágoras, vamos procurar explicar-vo-lo. Não é verdade que isso vos acontece, por exemplo, no caso frequente em que vos deixais vencer pelo comer, pelo beber, pelo amor, que são coisas agradáveis? Podeis saber que estas coisas são más e nem por isso deixais de as fazer.' Eles concordariam. Em seguida, tu e eu perguntar-lhe-íamos: 'Por que julgais más estas coisas? É porque vos proporcionam um prazer de momento e por serem agradáveis ou porque acarretam, como resultado futuro, a doença, a pobreza e outros males do mesmo gênero? Se não ocasionassem, no futuro, nenhum desses males e só causassem prazer, qualquer que fosse o juízo que se fizesse da causa e da maneira, seriam ainda más?' Poderemos supor, Protágoras, que eles nos dariam outra resposta a não ser esta?: 'Não é pelo prazer que proporcionam de momento que elas são más, é por causa das suas conseqüências, as doenças e outros males'.<sup>155</sup>

Mas a disposição a não se deixar vencer pelo prazer é nomeada de "temperança". No entanto, não parece haver evidência textual para afirmar que Hobbes considera a temperança como a virtude mais elementar, a que serve para definir todas as demais, por estar incluída na definição das virtudes cardiais. Vou sugerir sucintamente - apenas como uma hipótese a ser confirmada ou não por um exame mais minucioso da obra de Hobbes - o modo como estas últimas poderiam estar conectadas:

<sup>154</sup> Esta afirmação é feita por Strauss, que a desenvolve <sup>?</sup> em conexão com as teses da filosofia política de Hobbes; *Op. cit.*, pp. 15-18. Ver também *República*, 427e; e *De cive*, III, XXXII, 120-121.

<sup>155</sup> *Op. cit.*, p. 72 na ed. portuguesa, por mim ligeiramente alterada. É importante acentuar que um "bem real" é um bem desejado por um indivíduo que sabe (julga com boas razões) tratar-se de um meio efetivo na causação do resultado ou fim desejado. Tal bem continua sendo definido com relação a um fim desejado por um indivíduo. Não é um bem absoluto, algo como um fim desejável em si mesmo, mas sempre uma ação cuja realização é o que o indivíduo considera, na situação em que se encontra, ser prudente querer.

i) *coragem* é a disposição para manter-se determinado a fazer o que deliberadamente se descobre ser racional fazer;

ii) *justiça*, a disposição a desejar fazer o que as leis da natureza prescrevem como *correto* (*righteous*, na ed. inglesa; *juste*, na francesa) fazer;

iii) *prudência*, a disposição para avaliar uma ação como boa na medida que contribui para prolongar a própria vida, porque quem é prudente deseja fazer o que melhor contribui para prolongá-la (e, evitar fazer o que tende a abreviá-la).

Cabe sublinhar ainda um ponto. As condições de verdade da afirmação “uma pessoa deseja *irracionalmente* fazer  $x$ ” são “tal pessoa deseja fazer  $x$ ” e “ $x$  é - considerando suas conseqüências para tal pessoa - uma ação má”. O caso mais revelador não é o do engano (ou mesmo do auto-engano), mas justamente aquele em que a pessoa sabe que  $x$  é uma ação má para ela. Neste caso, qual o sentido de dizer que a pessoa escolhe *irracionalmente* fazer  $x$ ? De que tipo de *inconsistência* ela pode ser acusada? A de estar tentando obter algo que lhe inviabiliza a satisfação de seu desejo de agir *prudentemente*.

Calcula a utilidade de referir o cálculo de utilidade e viver. calculo de utilidade de referir o cálculo de utilidade e viver.

4.13. A prudência racional. *ignora a suspensão da firmeza de direção para um fim escolhido. permite "racional-mente" considerar o free rider como irracional.*

A faculdade da *reta razão* está sendo tomada aqui como o elemento racional da *prudência racional*: dizemos que um agente *prudente* (podemos omitir o uso explícito de "racional") busca a maximização da satisfação dos seus interesses a longo prazo.<sup>156</sup>

Usando uma terminologia apresentada por Jon Elster, podemos dizer que *agentes prudentes* são *global maximizers* e não *local maximizers*, isto é, eles sempre escolhem - entre alternativas de ação - aquela cujo conjunto de resultados produz a maximização dos seus desejos a longo prazo (e não apenas no tempo próximo);<sup>157</sup> neste caso, o cálculo da utilidade de uma ação - operado pela *reta razão* - inclui um número suficiente de conseqüências para o agente antever se a ação conduz ou não ao fim desejado. Por isso, Hobbes inclui, entre as capacidades que compõem a prudência, uma imaginação rápida, guiada por uma "firmeza de direção para um fim escolhido", acompanhada de juízo e discrição.<sup>158</sup>

Agentes prudentes estão dispostos a abrir mão de parte de sua satisfação no presente se e somente se (i.e., por definição) isto implicar numa maior satisfação futura, quer dizer, num maior ganho no futuro do que a perda ou sacrifício que lhe é exigido no presente. Ou ainda, numa imagem proposta por Sidgwick, se considerarmos um "todo" constituído de todos os desejos presentes e futuros de uma pessoa, se ela é racionalmente prudente, o que lhe interessa fundamentalmente é a satisfação dos seus desejos maximizada nesse "todo". Portanto, a capacidade de agir racionalmente pressupõe, no caso da prudência, a capacidade de ser motivado e valorizar *no presente* as conseqüências de longo prazo.<sup>159</sup>

<sup>156</sup> A caracterização da *prudência* como uma faculdade cuja operação é um cálculo dos meios de realização do auto-interesse de longo prazo é a mais aceita. Ver, por exemplo, Mackie, *Ethics: inventing right and wrong*, p. 119. Segundo Ezequiel de Olaso, "*recta ratio*" tem três significados na teoria hobbesiana: "(a) representação de uma norma que há de reger a conduta; (b) *prudentia* ou  $\phi\rho\nu\nu\eta\sigma\iota\zeta$ ; e (c) 'lei da natureza' ou 'lei natural'; neste último caso foi forte a influência dos estóicos". *Thomas Hobbes y la recta razón*, p. 31.

<sup>157</sup> Há pelo menos duas indicações textuais que podem apoiar esta interpretação: "(...) en l'état de guerre, personne ne peut être assuré de sa conservation, ni espérer d'atteindre à une bien longue mesure de vie"; *De cive*, I, XVI, 88; e "[numa situação onde] não há sociedade, ... a vida do homem é solitária, pobre, sórdida, embrutecida e curta"; *Leviathan*, 13, 62.

<sup>158</sup> Cf. *Leviathan*, 8, 32-33.

<sup>159</sup> Cf. *Leviathan*, 15, 72: "; (...) jamais se poderá dizer que [são] contra a razão [as ações que] tendem para [o] benefício próprio" e "as ações mais razoáveis são as que melhor conduzem [ao benefício próprio]"; cf. também *De cive*, III, XXXI, 120: "(...) as coisas presentes caem sob os



"Prudência racional" é o nome da capacidade de avaliar uma ação não segundo o padrão do "apetite presente", isto é, segundo a intensidade do desejo que possamos ter por alguma de suas conseqüências imediatas, mas sim conforme o *padrão* estabelecido pelo cálculo do valor das conseqüências futuras - ou, para usar uma expressão hobbesiana, as conseqüências que são óbvias apenas à razão<sup>160</sup> - na satisfação do auto-interesse. "Auto-interesse" tem aqui o significado atribuído pelo Bispo Butler: "razoável amor de si próprio".<sup>161</sup> Este termo tem como parte essencial do seu significado a maneira reflexiva e calculada pela qual alguém busca a satisfação de suas aspirações, sejam elas representadas pela busca de *prazeres sensuais* ou *prazeres mentais* (estes sendo ou bem a *glória* ou algo redutível a ela).<sup>162</sup>

A afirmação de que sempre são racionais as ações que promovem o benefício próprio é uma versão fraca do chamado *egoísmo racional*. E esta é a doutrina adotada por Hobbes, o que se pode comprovar pelo exame das citações recém feitas.<sup>163</sup> Na verdade, há uma sutileza na versão hobbesiana do egoísmo racional: "benefício próprio" deve ser interpretado como "o que é pessoalmente mais vantajoso se o cálculo das vantagens também pondera as conseqüências não imediatas das ações". Aceitando-se esta interpretação, *prudência racional* e *egoísmo racional na sua versão fraca* significam a mesma coisa: nem sempre as ações que melhor conduzem ao benefício próprio são as imediatamente mais vantajosas.

*A rigor, não parece haver em Hobbes preferência temporal — o importante é a quantidade total de satisfação que, em princípio, é função da duração; refere-se, tb., como fonte de unidades de prazer ou satisfação, um por fim de si e tempo.*

sentidos, mas as futuras só são concebidas pelo raciocínio" e I, VII, 83: "(...) nada é feito contra o uso da reta razão, quando a conservação própria é buscada por todos os tipos de meios."

<sup>160</sup> Cf. *De cive*, III, XXXI. Ver a tradução de Renato Janine Ribeiro.

<sup>161</sup> Cf. Albert Hirschman, *The passions and the interests*; p. 39 na trad. brasileira. É importante assinalar - o que faz toda a diferença - que "razoável" tem um significado normativo mais forte para Butler: os princípios de ação ditados pela razão e aqueles ditados pela benevolência (que expressam um interesse direto na promoção do bem-estar dos outros, interesse da "Consciência" ou "sentido do dever"), quando aplicados a diferentes situações, resultam em prescrições das mesmas ações. Ver Sidgwick, *op. cit.*, principalmente pp.93-94 e 172-173. Para Kant, a prudência é apenas *amor de si mesmo* esclarecido e sagaz na escolha de meios para "alcançar uma vantagem pessoal durável", o que a caracteriza como fonte de máximas genuinamente morais. Cf. *Fundamentação da metafísica dos costumes, segunda secção*, BA 41, 44.

<sup>162</sup> Cf. a seguinte passagem: "... todo prazer mental ou é a glória (que consiste em ter boa opinião de si mesmo), ou termina se referindo à glória no final". *De cive*, I, II na tradução brasileira já indicada.

<sup>163</sup> Cf. definição dada por Kurt Baier, *op. cit.*, p. 201.

[4.14. O objeto da "filosofia moral".] *alt.*

Segundo Hobbes, o desacordo entre as pessoas sobre o que é correto fazer em circunstâncias particulares é algo que não se pode eliminar. Isto não é possível porque a razão, ao cumprir um papel meramente instrumental nas deliberações, não pode servir de árbitro capaz de estabelecer com *imparcialidade* qual interesse em conflito deve ser priorizado, uma vez que os retos raciocínios das partes em controvérsia estão a serviço de interesses inevitavelmente parciais. A *parcialidade* aqui pressuposta como inevitável é resultado da própria ordenação das preferências individuais, que formam "o ponto de vista" a partir do qual as leis da natureza seriam interpretadas em circunstâncias particulares.<sup>164</sup>

Isto porém não implica que, como no caso da aritmética, a *filosofia moral* não seja "infalível e certa", uma vez que há também - e este é o principal objeto do último capítulo desta dissertação - a possibilidade de tal cálculo ser feito usando-se, segundo Hobbes, tão somente definições e termos gerais. Isto significa que as conclusões serão válidas para *tipos genéricos de ações*<sup>165</sup> estratégicas e suas conseqüências. As definições e termos gerais são usadas para formar proposições genéricas ("princípios verdadeiros" é o termo usado por Hobbes) que servem de hipóteses sobre situações-tipo de interação. Estas hipóteses descrevem um "jogo": há indivíduos - motivados de uma certa maneira e de posse de um conjunto de informações sobre a situação em que se encontram - que devem fazer decisões que, em maior ou menor intensidade, determinam o resultado, isto é, um tipo de recompensa ou punição para cada um que participa. Neste caso, as conclusões dos raciocínios podem ser consideradas proposições universais e necessárias (isto é, conclusões derivadas dedutivamente das hipóteses iniciais) a respeito do papel de diferentes estratégias usadas pelos jogadores na determinação do "resultado do jogo".

Tais conclusões, quando válidas para um "jogo" complexo cuja descrição completa é o que Hobbes chama de "estado de natureza", são chamadas *teoremas da razão*. Este jogo pertence, segundo os teóricos da escolha racional, a uma família de diferentes

<sup>164</sup> Podemos lembrar aqui um aforismo de Bacon, de quem Hobbes foi secretário na juventude: "o intelecto humano não é luz pura, pois recebe influência da vontade e dos afetos, donde se poder gerar a ciência que se quer. *Pois o homem se inclina a ter por verdade o que prefere.*" *Novum organum*, af. 49; grifo meu. A geração da "ciência que se quer" poderia significar, no caso da teoria de Hobbes, a criação de um conjunto de regras parciais (servindo a vontades particulares) para a aplicação das leis da natureza.

<sup>165</sup> O termo hobbesiano é "ações universais". Ver a *Epístola dedicatória no De cive*.

modelos para analisar "princípios morais como soluções de problemas que seriam causados pela interação irrestrita de indivíduos auto-interessados e racionais".<sup>166</sup>

[A *filosofia moral* constitui-se no sistema de regras gerais sobre o que é *bom* e *mau*, e conseqüências da sua aplicação (todas elas estabelecidas com certeza, isto é, <sup>alt.</sup> dedutivamente a partir de princípios verdadeiros), "na conservação e na sociedade humana".]<sup>167</sup> No entanto, esta certeza não pode ser obtida no caso de ações particulares, porque não há nenhum "cômputo", realizável por um indivíduo, para derivar, de enunciados universais (regras gerais ou "teoremas da razão") - válidos para todas as circunstâncias -, conseqüências particulares.

Esta *certeza* (propriedade de enunciados deduzidos validamente de axiomas, que são *auto-evidentes*) também não pode ser estabelecida por consenso, simplesmente porque, para um grupo de indivíduos, o fato de o resultado do cômputo ter sido o mesmo, não garante, por si só, a veracidade de tal resultado. Isto porque, podemos conceber a possibilidade de um acordo sobre certos tópicos mesmo entre homens insanos, usando métodos pouco racionais para estabelecer os seus termos, o que contraria a idéia básica de ciência como um conjunto de crenças estabelecidas por raciocínios indutivos corretos e dedutivos válidos.<sup>168</sup>

#### 4.15. A necessidade do acordo sobre um "valor comum".

Segundo a concepção hobbesiana de racionalidade, a razão prática opera instrumentalmente sobre um conjunto dado de valores individuais, isto é, calcula e prescreve o que deve ser feito para melhor realizar tal sistema de valores. Com isso, um agente pode ser dito ter sempre razão de fazer aquilo que maximiza a promoção dos seus próprios interesses. No entanto, o que Hobbes procura mostrar empregando a ficção do *estado de natureza*, é necessário também pressupor que os agentes compartilhem um valor,

<sup>166</sup> Heap, *The theory of choice*, p. ix.

<sup>167</sup> Cf. *Leviathan*, 15, 79.

<sup>168</sup> Ver Bacon, *op. cit.*, af. 27. Para uma apresentação elementar da importante distinção entre *validade* de argumentos dedutivos e *correção* de argumentos indutivos, ver Wesley Salmon, *Lógica*, pp. 76-78.

um valor de *todos* (*valor comum*), porque sem ele é impossível para cada agente realizar aquilo que mais deseja, isto é, a *própria* preservação.

Isso, todavia, não se constitui em um contra-exemplo da tese da relatividade do valor. Dizer que algo é um *valor comum* não implica em necessariamente comprometer-se com a afirmação da *objetividade* de tal valor, porque um valor comum pode ser um *valor instrumental*, isto é, um fim que não é desejável intrinsecamente por todo ser humano, mas só é desejável por cada um por ser um meio necessário para a realização de sua *própria* felicidade. Assim, dizer que, se *A* é um ser humano, “*x* tem valor instrumental para *A*” não significa mais que dizer que “*x* é um meio que necessária e positivamente promove os *planos racionais* de *A* para realizar a sua *própria* felicidade”.

A guerra perpétua, diz Hobbes, é inadequada à conservação dos indivíduos. “Guerra perpétua” descreve uma condição na qual o padrão violento da interação entre os homens é causa dos males sofridos por cada um. Cada ação violenta (ou declaração de intenção ou mesmo presumível intenção de assim agir) que contribui para gerar o que chamamos de “estado de guerra” é um meio para más conseqüências.

Ações más são, por definição, aquelas que causam dano. Mas a que tipo de dano se está referindo? Se o dano e o benefício causados por uma certa ação forem somente medidos pela satisfação de “apetites presentes” (i. e., se forem avaliados apenas como “bens presentes”), ações como por exemplo *oprimir* e *arruinar* não serão julgadas valorativamente de acordo com o “lado” ocupado por cada um que julga: de oprimido ou opressor, de vencedor ou vencido?

Ainda assim, deve ser possível um acordo quanto aos bens futuros resultantes de ações presentes. Este acordo pressupõe uma capacidade de antever um mal futuro, isto é, a capacidade de sentir medo, de ter uma aversão que resulta da crença de uma conseqüência danosa futura de uma ação que se faz no presente.<sup>169</sup> Sob estas condições, é racional sentir medo. E é precisamente na pressuposição desta paixão, capaz de motivar mesmo os mais socialmente perversos e avessos à cooperação, que está assentada a “base realista” da ética de Hobbes.

<sup>169</sup> Cf. as definições complementares de medo, apresentadas por Hobbes no *De cive*, I, II, última nota e *Leviathan*, 6, 25.

#### 4.16. A paz como "valor comum".

Que todos os homens desejam a sua própria conservação pode significar duas coisas: que desejam somente (e por isso agem para obter) o que lhes parece o mais agradável de imediatamente obter, por preferirem gananciosamente um bem presente a um bem futuro, ou - sendo capazes de ponderar os benefícios e danos não imediatos do que decidem fazer e de preferir um maior bem futuro a um menor bem presente - que desejam racionalmente (o que, pela definição de razão, exclui desejar graças a alguma inclinação ou apetite *imediate* ou *não deliberado*; ou como Hobbes diria, um tipo de aptidão ou disposição natural) todos os meios que "tendem a sua própria conservação".<sup>170</sup>

A paz, entendida como o conjunto das ações cooperativas que dão estabilidade às sociedades - as "sociedades civis", diz Hobbes, definem-se não como meras reuniões de indivíduos, mas como um conjunto de ações feitas para cumprir com obrigações (*bonds*) recíprocas direta ou indiretamente acordadas; este último caso referindo-se às obrigações mais abstratas cuja racionalidade da aceitação individual voluntária é justamente o principal objeto de prova das teorias do contrato social -, é um meio para cada indivíduo poder satisfazer seus próprios desejos.

Adotando uma perspectiva que integra as considerações hobbesianas anteriores,<sup>171</sup> pode-se dizer, num estilo neo-hobbesiano, que não se trata de entender a paz (o *valor comum* assumido por Hobbes) como algo intrinsecamente valioso para todo e qualquer ser racional, mas somente como um meio racional (mais eficiente) para - que deve ser perseguido (porque aceito como racional) por - "indivíduos auto-interessados e seculares que enfrentam os [i. e., dão-se conta dos] conflitos [conseqüentes ao] egoísmo crasso [*naked egoism*]"<sup>172</sup>.

<sup>170</sup> A distinção conceitual que apresento neste parágrafo está presente na seguinte passagem do comentário clássico de Leo Strauss:

"The striving after power may thus be rational as well as irrational. Only the irrational striving after power, which is found more frequently than the rational striving, is to be taken as natural human appetite. For the rational striving after power rests on already reflection and is for that very reason not natural, i. e. not innate, not in existence prior to all external motivations, to all experience and education." *The political philosophy of Hobbes*, p. 10.

<sup>171</sup> Conferir a letra hobbesina em *De cive*, I, II, IX, XII e XIII; e III, XXXI e XXXII.

<sup>172</sup> David Gauthier, *Why ought one obey God?* in *Moral dealing*, p. 35. Neste sentido, o tipo de justificação da moralidade que Hobbes adota pode ser denominada como *justificação teleológica limitada*. Isto quer dizer, nem sempre é racional - para o tipo de pessoa que Hobbes pressupõe situada nas condições interativas que também pressupõe - fazer o que maximiza os seus próprios

As “leis da natureza” são preceitos da razão que expressam o que deve ser feito para que a paz seja realizada. E, ainda que expressando verdades, tais preceitos são evidentes apenas para aqueles que “sem parcialidade ou paixão fazem uso de sua razão natural”. É preciso pressupor uma dada psicologia, a presença de certas “qualidades que predisõem os homens para a paz e a obediência”,<sup>173</sup> sem a qual tais “preceitos” não serão *verdades indiferentes* (para usar uma expressão de Hannah Arendt em *Verdade e Política*), aceitáveis por todos os homens por não se opor aos interesses e prazeres de nenhum deles. Acontece que alguns homens, mesmo tendo segurança para respeitar às “leis da natureza”, deixam de respeitá-las porque *sim se preocupam* (“care” é o verbo é usado por Hobbes) com as verdades por elas expressas: elas contrariam os seus interesses e prazeres, principalmente os associados à ambição, à cobiça e ao lucro.<sup>174</sup>

#### 4.17. A concepção hobbesiana da moralidade.

Resta ainda esclarecer um outro problema. Se a aceitação de algum valor compartilhado implicar, para cada agente, a adoção de uma outra perspectiva organizadora de seus próprios valores - na qual a hierarquia de preferências não permanece idêntica à anterior, produzindo, por isso mesmo, uma alteração do que é racional fazer - que tipo de agente (disposições, motivações, interesses, crenças, etc.) seria capaz de aceitá-lo racionalmente? No próximo capítulo, algumas considerações servem para elucidar este problema.<sup>175</sup>

Por enquanto, podemos ver tais dificuldades desde um ponto de vista ligeiramente diferente: como as exigências da moralidade, concebidas essencialmente como razões universais - isto é, razões que indicam o que é melhor para cada um e todos os homens, não como comportamento instrumental, quer dizer, independentemente de quais

---

valores, independentemente de quais sejam. Precisamente, se a busca destes valores for incompatível com a produção da paz - valor comum - então não é racional manter tal busca. Para uma discussão das relações teóricas possíveis entre *moralidade* e *valor* e o modo como elas afetam a concepção de *justificação moral*, ver Gaus, *op. cit.*, pp. 13-22.

<sup>173</sup> *Leviathan*, 26, 138 e 143 respectivamente.

<sup>174</sup> Cf. *Leviathan*, 11, 50; 17, 85; e revisão e conclusão, 396.

<sup>175</sup> Ver principalmente 5.2 e 5.17.

sejam os seus desejos e preferências ligados a resultados - poderão ser derivadas de uma teoria do valor subjetivista?

Esta pergunta, no entanto, pressupõe implicitamente uma concepção de moralidade que não é a de Hobbes. Que fique feita a seguinte observação: a dificuldade apontada é uma aporia insolúvel se concebemos a moralidade como expressão (e busca) de uma ordem transcendental ("um mundo como um bem supremo", diz Kant), o ponto de vista moral como basicamente diferente do auto-interessado, as motivações do agente moral como isentas da inclusão de qualquer fim particular (o que pressupõe a *renúncia*, pelo homem que quer agir moralmente, da busca do seu "fim natural", ou seja, "sua própria felicidade") e envolvendo uma aplicação racional de um princípio imparcial que limita a vontade a obedecer incondicionalmente - i. e., a motivação para a obediência não pode ser outra senão a do respeito de um ser racional ao *dever* - a máximas completamente universalizáveis.

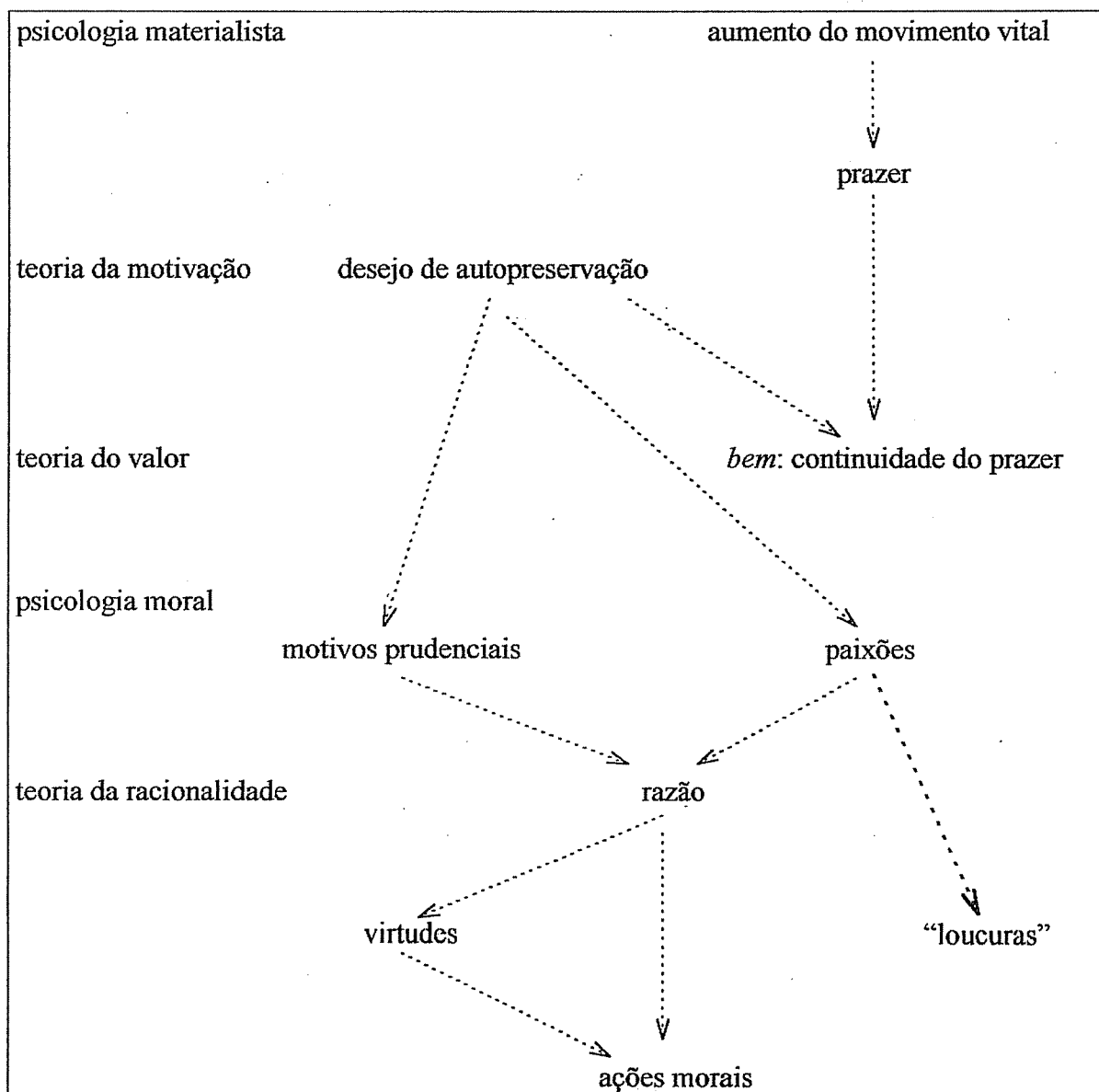
Contudo, na concepção hobbesiana de moralidade, a que guia a construção da sua teoria ética, a moralidade é o produto da capacidade racional de indivíduos auto-interessados de impor sobre si mesmos restrições limitando a busca cegamente competitiva da sua própria felicidade; tais restrições são impostas por meio da adesão de cada indivíduo às *leis da natureza* (que também são máximas *universalizáveis*, ainda que não completamente), motivado a abrir mão de uma máxima satisfação de seus interesses imediatos sempre que tal adesão resultar num maior benefício futuro para si, resultado este que está condicionado a uma *adesão multilateral*, quer dizer, resultado que supõe ser, sob certas condições, a adesão individual condição necessária e suficiente para a adesão coletiva.<sup>176</sup>

<sup>176</sup> Antecipando um pouco uma discussão a ser feita mais adiante, pode-se sugerir que a disputa mais fundamental é a que gira em torno da caracterização da racionalidade prática - ou, a disputa filosófica sobre a possibilidade da razão ser usada para orientar decisões sobre o que fazer. Pode-se resumir o conflito teórico em questão dizendo que há basicamente três visões filosóficas conflitantes:

"Ser racional na prática, um grupo afirma, é agir baseado em cálculos de custos e benefícios, para si mesmo, de todos os cursos de ação possíveis e suas conseqüências. Ser racional na prática, afirma um grupo contrário, é agir sob restrições tais que qualquer pessoa racional, capaz de uma imparcialidade que não concede nenhum privilégio particular aos interesses próprios, concordaria que devem ser impostas. Ser racional na prática, diz um terceiro grupo, é agir de maneira a alcançar o último e verdadeiro bem dos seres humanos" MacIntyre, *op. cit.*; p. 12 na ed. brasileira.

MacIntyre pretende, neste livro, localizar diferentes filósofos - Platão, Aristóteles, São Tomás de Aquino e Hume, entre outros - como participantes de algum destes grupos ou tradições. Como

*Arquitetônica da teoria ética de Hobbes.*



exemplos mais apropriados para a presente discussão, podemos exemplificá-los respectivamente por Hobbes, Kant e Aristóteles.



## 5. JUSTIFICAÇÃO DAS AÇÕES MORAIS

"Qual teoria da moralidade pode alguma vez servir a qualquer propósito útil, a não ser conseguindo mostrar que todos os deveres que prescreve são também do interesse real de cada indivíduo?"

David Gauthier, *Morals by agreement*

"Uma sociedade constituída de egoístas irrefreados se espatifaria em pedaços"

Joan Robinson, *Economic philosophy*

### 5.1. A "condição natural da humanidade".

É bem conhecida a descrição que Hobbes faz da "condição natural da humanidade". Trata-se de uma hipótese teórica sobre a totalidade das relações humanas na ausência de um *poder comum* capaz de manter os homens - pelo receio que lhes impõe - em respeito, e isto envolve (i) garantir, a cada um, proteção contra todos os outros e (ii) governar as ações de cada um no sentido do *benefício comum*.<sup>177</sup> A caracterização desta condição como um "estado de guerra de todos contra todos" não deve ser tomada literalmente. Ela visa chamar a atenção para traços que são compartilhados por "tempos de guerra" e por "tempos de paz": o comum é que "os homens vivem sem outra segurança senão a que lhes pode ser oferecida por sua própria força e sua própria invenção".<sup>178</sup>

Ainda que não haja um acordo entre os comentadores quanto à validade do argumento hobbesiano para justificar a hipótese teórica de que, se são assumidas certas premissas sobre a natureza humana (incluindo entre elas certas premissas sobre determinadas relações de comparação entre diferentes indivíduos), o "estado de natureza" (definido pela natureza dos indivíduos e de relações entre eles) é um "estado de guerra de todos contra todos", parece haver acordo entre eles no que diz respeito ao valor desse

<sup>177</sup> Cf. *Leviathan*, 13, 62 e 63, e 17, 86. A natureza coercitiva deste "poder comum" será melhor apreciada mais adiante.

<sup>178</sup> *Leviathan*, 13, 62.

argumento para a compreensão de alguns problemas fundamentais relativos à interação humana.<sup>179</sup>

Primeiramente, se assumimos que, no que diz respeito à força e à invenção, os homens são por natureza iguais, enquanto subsistir o "estado de natureza", nenhum homem pode ter certeza de que sua vida estará garantida num momento futuro. Por mais próximo que seja tal momento, pois, dada a igualdade dos homens quanto a sua natureza física e mental, cada homem não pode esperar razoavelmente ter mais chance que os demais em conseguir atingir seus fins (que são potencialmente, direta ou indiretamente, objetos do desejo dos demais). Existe então esta igualdade crucial: no "estado de natureza", a sobrevivência de cada um é igualmente provável.

Em segundo lugar, dadas estas condições de incerteza, cada um, para atingir seus fins, toma como sua *estratégia dominante* - isto é, a estratégia que é, desde o seu ponto de vista, a melhor resposta a todo e qualquer movimento dos outros indivíduos com os quais virá a interagir - a chamada *antecipação* ou, como ficaria mais claro dizer, *violência antecipatória*: procurar antecipar-se aos possíveis ataques de outros e, pela força ou pela astúcia, subjugar-los ou mesmo destruí-los.<sup>180</sup>

Podemos imaginar que, no estado de natureza, ninguém se pergunta, ao decidir o que fazer, se algum bem-estar para os outros resultará do curso de ação escolhido. Trata-se de uma situação interativa possível, logicamente coerente. Mas ela não é necessariamente a descrição da interação que se dá unicamente entre agentes motivados pela cobiça (ou outra paixão da qual possa depender o chamado comportamento de *competidor*), pois basta que aqueles que interagem não saibam quem é quem, isto é, não tenham condições de distinguir quem se comportará exclusivamente por causa desta *motivação egoísta cega* (não ponderada) de quem tenderá a se comportar deste modo apenas como último recurso para garantir a sua sobrevivência. Assim, não é preciso tomar um egoísmo orientado por paixões como esta como uma *motivação universal*, basta introduzir restrições na informação que cada agente dispõe para tomar suas decisões sobre o que fazer, em dadas circunstâncias, para poder continuar vivendo.<sup>181</sup> O mais importante é

<sup>179</sup> Cf. Gregory Kavka, *Hobbes's war of all against all*, p. 291. Neste artigo, Kavka pretende mostrar por que esse argumento hobbesiano é, em última análise, inválido.

<sup>180</sup> Cf. *Leviathan*, 13, 61.

<sup>181</sup> Algumas outras paixões como a cobiça quanto às ações que geram: a ira, a vanglória, o ódio.

Uma  
versão  
educada  
do estado  
de natureza  
é o contrato  
social

perceber que atribuir a estas paixões o papel de motivação fundamental *de todos os homens* inviabiliza a busca racional da garantia e manutenção da paz (quer dizer, nem se instaura o contrato, nem se mantém a sociedade). Este ponto será melhor articulado mais adiante, por meio dos conceitos de *estabilidade racional e motivacional* das "leis da natureza".

Como vimos, uma tese radical da teoria do valor do egoísmo pode ser assim enunciada: a continuidade do *próprio* prazer é realizada, segundo cada indivíduo, por sistemas de valores próprios e divergentes, sistemas que guiam as escolhas, ao ingressar nas deliberações oferecendo hierarquias de preferências por ações, regras, finalidades, pessoas, objetos, etc. Disso se segue uma importante consequência: é precisamente este desacordo quanto aos valores entre os homens que está na base da explicação para o fato de "a condição natural da humanidade" ser - para usar um conceito de David Gauthier - uma "condição radicalmente insuficiente", isto é, uma condição que apresenta os seguintes traços definidores:

i) nela, as ações realizadas por cada indivíduo no esforço de manter, da melhor forma possível, o seu movimento vital, quando também realizadas similarmente pelos outros, terminam criando um padrão de interação (precisamente, a interação *entre inimigos*) que na realidade constitui-se num obstáculo à manutenção do movimento vital individual;

ii) nela, supondo que todos os indivíduos são iguais quanto às suas capacidades de infligir danos e suscetibilidades de sofrê-los (o que define a condição de *igualdade natural*), a diminuição total de movimento vital que cada um sofre é maior que o aumento total que pode obter infligindo danos aos demais (usando a linguagem econômica, na interação humana dominante no estado de natureza, cada indivíduo está impossibilitado de obter um resultado otimizado - i. e., ter o maior benefício compatível com os benefícios buscados pelos demais - pois, a taxa de acréscimo dos custos é superior relativamente à de acréscimo dos benefícios (os custos marginais excedem os benefícios); por isso pode-se dizer que a "condição natural da humanidade" é tal que nela os indivíduos encontram-se num estado pior do que estariam num estado de solidão, quer dizer, não interagindo (ou, necessitando interagir apenas minimamente) com os demais.<sup>182</sup>

No plano motivacional, esse estado de guerra total define-se fundamentalmente pela predominância de homens que estão dispostos (a natureza da guerra não é definida em termos de uma luta real, mas, da *disposição para lutar*) a buscar apenas a máxima satisfação imediata dos seus desejos (*local maximizers*, na terminologia antes sugerida). Movidos por

<sup>182</sup> Ver Gauthier, *Three against justice: the fool, the sensible knave, and the lydian shepherd in Moral dealing*, pp. 130-134.

um *apetite brutal* - a cobiça - eles disputam a conquista dos bens ao seu alcance usando de todos os meios. "Cobiça" não se refere aqui a um apetite particular - i. e., por algo determinado, a saber, a riqueza - mas sim a um modo de buscar a satisfação de diversos apetites particulares, a saber, a busca míope da satisfação (imediate) destes apetites.<sup>183</sup>

Apesar disso, por serem capazes de um raciocínio sobre a estratégia necessária para prolongar ao máximo a sua própria conservação - o raciocínio que não somente lhes capacita a agir visando satisfazer o *desejo de maximizar globalmente*, mas que também é regulado por este desejo<sup>184</sup> -, podem estar de acordo, sempre que realizam o cálculo corretamente, sobre o que deve ser minimamente feito para garantir, como consequência, a continuidade da satisfação dos seus desejos: a *paz* é o meio necessário para obterem os seus próprios fins.<sup>185</sup>

<sup>183</sup> Por isso, não se apercebem - o que é unicamente possível por meio de um processo de *deliberação ideal* (i. e., livre da influência de paixões inadequadas ou de cálculos insuficientes) - de que as ações que escolhem não são os *meios mais eficazes* para a sua própria conservação. Como consequência importante, mesmo aqueles que reconhecem (por meio de um cálculo correto) que trapacear, fraudar, pilhar, etc., não são as ações mais eficazes para obterem uma maximização da satisfação global dos seus desejos, não deixam de levar a cabo ações deste tipo, simplesmente porque seria *irracional* de sua parte não prevêê-las por parte dos indivíduos pertencentes ao grupo dos "predadores". Ver discussão em 3.4..

<sup>184</sup> Para a caracterização das "cadeias de pensamentos" regulados continuamente por algum desejo ou desígnio, ver *Leviathan*, 3, 9.

<sup>185</sup> Gostaria de apontar pelo menos duas interpretações do argumento hobbesiano usado para provar a proposição de que o estado de natureza é um estado de guerra de todos contra todos. Gregory Kavka não vê nenhuma ambigüidade nas premissas deste argumento, e, após listá-las, procura indicar por que razões ele não pode ser aceito como válido. Ver o artigo *Hobbes's war of all against all*.

Jean Hampton, ao contrário, reconstrói diferentes versões daquela prova. Algumas, sustenta esta intérprete, de fato não são suficientes para produzir o resultado desejado. Uma delas, porém, é um argumento válido. Na análise desta autora, fica mais clara a complexidade e variedade dos traços psicológicos dos agentes hobbesianos. As versões insuficientes são analisadas no capítulo intitulado "*What is the cause of conflict in the state of nature?*" do livro *Hobbes and the social contract tradition*; o argumento válido, denominado pela autora de "*the shortsightedness account of conflict*", é analisado no capítulo 3 do mesmo livro.

## 5.2. O que um "egoísta racional" busca?

Como já foi indicado, o modelo de racionalidade adotado por Hobbes pode ser descrito pela expressão "a razão prática como instrumento da satisfação de desejos". Mas não de quaisquer desejos. A limitação na satisfação de desejos é introduzida exatamente pelo desejo de maximizar globalmente, prolongar ao máximo a própria conservação. Tal desejo funciona como *razão interna* para justificar a adesão a normas que limitam a busca da maximização da utilidade por parte dos agentes. Isto é necessário para torná-los agentes *maximizadores cooperativos* (termo usado por L. Mulholland). O *desejo de maximizar globalmente* também é um motivo egoísta, isto é, trata-se do que é necessário para a obtenção de vantagens individuais do modo mais eficiente. No caso, os agentes têm razão para aderir a normas que limitam a sua própria satisfação porque é justamente por meio de uma tal limitação no comportamento que o desejo de prolongar ao máximo a própria conservação será satisfeito.<sup>186</sup>

O desejo de maximizar globalmente é o desejo que torna possível a prudência. Com isto, o tipo de egoísmo aqui caracterizado tem a prudência como um dos seus traços definidores. Trata-se do tipo denominado por Sidgwick de "egoísmo racional" (ou "prudência racional"). O termo "prudência" é comumente tomado como sinônimo de "auto-interesse". Mas isso não é adequado, pois também é possível um comportamento guiado pelo "auto-interesse crasso", que se caracteriza por uma perseguição irrefreada do que *eu* (um e só um determinado indivíduo) quero, sem considerar a minha relação, por meio de minhas ações, com "as vidas de outras pessoas", isto é, sem que quaisquer desejos, reivindicações, carências, etc. destas pessoas constem como parte do cálculo racional que realizo para escolher *o que fazer* em diferentes circunstâncias.<sup>187</sup>

<sup>186</sup> A partir de uma distinção feita por Bernard Williams em *Internal and external reasons*, artigo incluído na coletânea *Moral luck*.

<sup>187</sup> Williams adequadamente distingue esta forma mais "rasa" de egoísmo do chamado "egoísmo ético" (uma expressão que considera enganadora). Este se define como a teoria segundo a qual "cada pessoa deve perseguir o seu próprio auto-interesse". E se distingue, continua Williams, do tipo "raso" por ser "uma posição reflexiva e [assumir] uma visão genérica acerca dos interesses das pessoas". Ver *Ethics and the limits of philosophy*, pp. 11-12. O termo "auto-interesse crasso" é usado pelo economista Eduardo Giannetti da Fonseca para referir-se a um tipo de princípio de conduta (e a motivação correspondente) definido pela conjunção de egoísmo (o mesmo que "egoísmo raso") e "oportunismo", este definido como o comportamento daqueles que "jamais [perdem] uma chance sequer de tirar proveito em benefício próprio da violação de normas de convivência social". *Vícios privados, benefícios públicos?*, p. 143.

Além disso, o elemento central de uma ação prudente é precisamente o fato de ter sido adotada com base no princípio segundo o qual as conseqüências futuras da ação devem ser ponderadas adequadamente, isto é, não podem perder "peso" (não sendo assim avaliadas como "piores", isto é, como "menos valiosas") simplesmente por não representarem uma satisfação imediata de desejos e interesses. Para precisar o sentido que estou atribuindo ao termo, emprego uma colocação de Thomas Nagel, dizendo que as razões para alguém agir prudencialmente "originam-se de um desejo presente de satisfazer [seus] desejos e interesses futuros (ou de um desejo de ter uma satisfação a longo prazo de todos os [seus] interesses, presentes e futuros)".<sup>188</sup> Assim, diz-se que alguém é prudente quando é capaz de renunciar à satisfação presente de alguns dos seus desejos sempre que isso implicar na obtenção de uma maior satisfação futura, de desejos que tem no presente ou que terá no futuro, mesmo que não os tenha no presente.

Assim interpretado, o princípio da prudência "está conectado com uma concepção da situação presente de alguém como meramente uma etapa em uma vida extensa temporalmente". Esta concepção "surge da capacidade humana de ver o presente simultaneamente como 'agora' e como um momento particular [qualquer]".<sup>189</sup> Supõe-se um agente capaz de ver o presente como meramente um momento entre outros sem especificação temporal. Isto é, dito aproximadamente, capaz de aplicar um princípio de imparcialidade sobre o conjunto dos momentos da sua vida: não constitui uma razão suficiente para preferir um *bem presente* (isto é, a satisfação imediata de um desejo que se tem 'agora') a um *bem futuro* (i. e., a satisfação futura de um desejo que se tem 'agora' ou que se terá no futuro) o simples fato de ser um bem presente (supondo-se desejos com a mesma intensidade e igualmente prováveis de serem satisfeitos). Temos então a idéia de um indivíduo que se concebe como durando no decorrer do tempo, capaz de 'projetar-se' no futuro e conceber a *si mesmo* (i. e., conceber-se como um *eu* que permanece) como portador de estados futuros de consciência.

Duco esse que não há, em Hobbes, este cuidado com uma ponderação de fins longos e sustentados de satisfações presentes e futuras. O que se faz é saber a qual espécie de vida optar: o atalho

<sup>188</sup> *The possibility of altruism*, p. 36.

<sup>189</sup> *Ibid.*, p. 19.

felo temer de morte, pela preservação de vida que é a condição de possibilidade de todos e para a satisfação presente e futura. Isto é, por uma avaliação de tudo por meio de desejos, não apenas a possibilidade, necessariamente incerta, de maior satisfação futura.

### 5.3. As "ações estratégicas" necessárias para a paz.

Como é possível determinar o que é necessário fazer para garantir paz, que é, por sua vez, um meio necessário para todos os indivíduos (e não um fim comum) conseguirem realizar os seus fins últimos individuais, isto é, a continuidade, para cada um, de sua própria vida? Fazendo uso de raciocínios corretos cujas conclusões são generalizações sobre *ações estratégicas*. São estratégicas porque (1) são pensadas como aplicáveis a uma série de interações de cada agente com os demais e (2) são escolhidas em circunstâncias nas quais os demais também as levam em conta nas suas próprias escolhas e, portanto, impõem a cada agente fazer um cálculo no qual a "reação" possível dos outros é igualmente considerada como uma variável.

Como vimos, a ação estratégica dominante no estado de natureza é a *antecipação*. Cada agente adota, como estratégia, tal ação porque calcula erroneamente que ela é a que lhe trará o maior benefício. Há uma série de falsas hipóteses que conduzem-no a essa conclusão: p. ex., ele crê ser mais astuto e mais forte que os outros; também crê que não interagirá mais de uma vez com o mesmo indivíduo ou ainda, que o seu comportamento de rapina não será memorizado pelos que futuramente vão de novo interagir com ele.

Neste ponto, é importante recordar uma causa importante das "falhas da racionalidade" nas deliberações. O que é comum às várias crenças recém apontadas é o fato de não existirem boas evidências para sustentá-las. Ou melhor, as evidências que porventura pudessem apoiá-las nem sequer são buscadas. Eis um caso onde, por interferência direta dos desejos, as crenças são escolhidas sem um exame racional das evidências. O agente passa a crer que algo é o caso simplesmente porque ele deseja que seja assim. Hobbes aponta um modo pelo qual o agente consegue enganar a si mesmo, porque descuida do exame daquelas evidências que podem contrariar o que ele deseja que seja verdadeiro:

*"Pois a natureza dos homens é tal que, embora sejam capazes de reconhecer em muitos outros maior inteligência, maior eloquência ou maior saber, dificilmente acreditam que haja muitos tão sábios como eles próprios; porque vêem sua própria sabedoria bem de perto, e a dos outros homens à distância"<sup>190</sup>*

Acontece que, se a maior parte dos indivíduos raciocina do mesmo modo, aquela que individualmente parecia ser a melhor estratégia para agentes que visam maximizar a quantidade de prazer que experimentam -, após ser adotada coletivamente (incluindo aqueles indivíduos que a adotaram como último recurso), torna-se uma *estratégia irracional* porque mais danosa que benéfica para cada um. Na "condição natural" recém descrita, nenhuma *situação contratual* (onde promessas mútuas caracterizam um pacto para a realização, de parte à parte contratante, de ações visando vantagens mútuas) pode chegar a concretizar-se, simplesmente porque ninguém pode nem mesmo aceitar uma promessa.

---

<sup>190</sup> *Leviathan*, 13, 61. Também é importante apresentar neste ponto algumas considerações feitas por Anatol Rapoport sobre a racionalidade dos "bons jogadores" nos chamados "jogos de estratégia" (o xadrez é um exemplo deles). Este autor sublinha acertadamente que "presumir que o adversário cometerá o erro que desejamos que cometa não é em geral realista, pelo menos na ausência de provas da capacidade inferior do outro".

Segundo Rapoport, a análise destes jogos propicia um bom modelo do comportamento racional de pessoas que estejam em situações cujos traços definidores gerais são:

- 1) a existência de conflitos de interesses;
- 2) a existência, em cada interação entre as pessoas, de um certo número de cursos de ação alternativos;
- 3) o fato de cada pessoa ter condições de avaliar as conseqüências de suas escolhas e que, para isso, leva em conta o fato crucial de que tais conseqüências não dependem unicamente das suas próprias escolhas, mas também dos cursos de ação adotados pelas outras pessoas com quem interage, sobre cuja escolha não tem controle. Cf. *Lutas, jogos e debates*, p. 88.



#### 5.4. A natureza do "estado de guerra".

Por isso, o "estado de guerra" pode ser caracterizado como um *estado de não-cooperação total*, isto é, um estado onde todas as ações são provavelmente não-cooperativas, como por exemplo, fraude, trapaça, invasão, subjugação, etc. Um estado de não-cooperação total não necessita ser um estado em que todos os homens estão constantemente lutando uns contra os outros (uma "luta real", como Hobbes o diz); basta que seja um estado no qual, baseando-se em acontecimentos recorrentes, cada um tem uma grande expectativa da (atribui uma grande probabilidade à) ocorrência de atos de violência, coerção, roubo e fraude.

Isto implica em dizer que "a vontade de travar batalha" pode ser confirmada pela experiência. Ela pode ser conhecida empiricamente quando antes se formula uma hipótese sobre a frequência de ocorrência daqueles tipos de atos a partir da observação da repetição ordenada de uma seqüência de eventos. Deste modo, assim como "a natureza do mau tempo" é definida por uma "tendência para chover" - confirmada pela alta frequência de ocorrência da sucessão representada pelo condicional "nuvens carregadas → chuva" -, a "natureza da guerra" é definida por uma disposição para a luta.

O argumento forte de Hobbes não pretende, no entanto, ser apenas probabilístico. Hobbes quer fazer valer o que chama, no cap. XIII do *Leviathan*, de uma "inferência feita a partir das paixões". Trata-se, portanto, de estabelecer conclusões certas sobre a totalidade das interações humanas na suposição de não haver um poder comum que mantenha os homens em respeitoso temor ("*keep them in awe*"). Portanto, as premissas deste argumento não podem ser apenas enunciados contingentes sobre disposições humanas.

Usando a terminologia da teoria dos jogos, o "estado de guerra" é uma situação de "conflito puro" ou de "luta", isto é:

a) ela está estruturada (i. e., dados o sistema de preferências dos agentes, o tipo de raciocínio que levam a cabo, as informações de que dispõem) de tal modo que cada agente, a cada interação com os outros - suposta como do tipo *tudo-ou-nada* - decide adotar a única estratégia que acredita capaz de lhe poupar da perda total: buscar a ação que lhe garanta conseguir imediatamente o máximo possível;

b) ela é uma situação na qual a combinação da adoção por cada um dos agentes da estratégia que considera a melhor para si mesmo produz um resultado que é pior para

todos do que o que seria produzido pela adoção de uma estratégia, digamos, *menos cobiçosa*.

### 5.5. O que significa: "algo é um meio necessário para produzir a paz"?

Certamente, "produzir a paz" não significa cessar momentaneamente a guerra, mas sim alterar a natureza mesma do "estado de guerra". E isto pode ser obtido se introduzimos uma *condição de possibilidade* de toda e qualquer *ação cooperativa* (os termos gerais usados por Hobbes para referir-se a este tipo de ação são "*mutual help*" e "*assistance*"): a garantia do cumprimento dos pactos. Portanto, a adesão mútua aos pactos feitos é, desde o ponto de vista da racionalidade do agente hobbesiano, a condição *sine qua non* da supressão da "condição de guerra". A regra geral ou *teorema da razão* que enuncia esta condição é a "terceira lei da natureza": que os homens cumpram os pactos que celebrarem. Esta "lei" é também "fonte" de uma importante virtude moral: a *justiça*. Age injustamente aquele indivíduo que viola um pacto ao qual previamente assentiu.<sup>191</sup>

Não estamos diante de uma mera *permissão* genérica,<sup>192</sup> mas sim de uma *obrigação moral*. A natureza mesma desta obrigação lança luz sobre o tipo peculiar de *justificação da moralidade* que está em jogo no projeto hobbesiano. Devemos agir do modo prescrito pelas regras morais pela simples razão de que isto é o que garante da melhor forma, para cada um de nós, a realização dos nossos interesses individuais. Esta é uma idéia completamente alheia a qualquer teoria ética altruísta (as mais conhecidas são a kantiana e o utilitarismo clássico) porque este tipo de teoria considera de algum modo espúria a idéia de uma *razão para agir* moralmente que seja *auto-interessada*.<sup>193</sup>

<sup>191</sup> Cf. *Leviathan*, 15, 71.

<sup>192</sup> Hobbes endossa a tese genérica denominada *racionalismo ético*, ao afirmar que "se [uma ação] não é contra a razão, não é contra a justiça". *Leviathan*, 15, 72. Baier define *racionalismo ético* como "a doutrina de que uma exigência ou recomendação moral é válida ou aceitável se o seu cumprimento é algo que está de acordo com a razão". *Op. cit.*, p. 201. A tese do racionalismo ético expressa uma importante conexão entre o que é *moralmente permissível* e a racionalidade prudencial. Todavia, ainda é somente uma conexão fraca, pois não estabelece nenhum conteúdo em particular para a *lei natural*.

<sup>193</sup> Kant, para citar apenas uma passagem, afirma que "nem todo fim é moral (não o é, por exemplo, o da felicidade pessoal), mas aquele que o é deve ser desinteressado" *Théorie et pratique*, p. 17; Ak. 279-280.

Uma afirmação de Thomas Nagel pode ser tomada como exemplificando com precisão o que é talvez o cerne das teorias éticas altruístas:

*"A obrigação moral é algo que cumpre um papel nas deliberações, e que tem influência em situações nas quais uma pessoa não poderia realizar uma ação caso considerasse somente o seu próprio benefício; enquanto a consideração de uma obrigação moral - ajudar outras pessoas, por exemplo - leva-a de qualquer modo a realizá-la. Nada que por princípio nunca conflitasse com o auto-interesse poderia ser denominado como obrigação moral"*<sup>194</sup>

Se atentamos diretamente para a forma dos enunciados das "leis da natureza", podemos deixar de interpretá-las como máximas para a preservação individual e tomá-las erroneamente como enunciados de "princípios para a preservação constante dos homens em multidões, baseados no interesse pela conservação da sociedade ou dos homens em geral",<sup>195</sup> pois podemos tomá-las por *imperativos hipotéticos* que prescrevem ações que contribuem para a realização da paz como meio para a sobrevivência da sociedade. As "leis da natureza" poderiam, assim, ser erroneamente consideradas como instanciações da seguinte forma geral (a "lei primeira e fundamental da natureza"):

"A paz deve ser buscada"<sup>196</sup>

Hobbes apresenta diversas vezes a razão pela qual devemos buscar a paz. Em uma delas, tal razão aparece na própria definição de "lei da natureza":

*"Uma lei da natureza (lex naturalis) é um preceito ou regra geral, estabelecido pela razão, mediante o qual se proíbe a um homem fazer tudo o que possa destruir sua vida ou privá-lo dos meios necessários para preservá-la, ou omitir aquilo que pense poder contribuir melhor para preservá-la"*<sup>197</sup>

<sup>194</sup> *Hobbes's concept of obligation*, p. 74; grifo meu.

<sup>195</sup> *Ibid.*, p. 71. A propósito, este é, segundo Nagel, um dos equívocos da obra que considera como "uma importante contribuição à nossa compreensão das idéias de Hobbes" (p. 68), o livro *The political philosophy of Hobbes* (1957), de Howard Warrender.

<sup>196</sup> Hobbes a enuncia como "procurar a paz, e segui-la", *Leviathan*, 14, 64. Não é necessário que um enunciado contendo o termo "deve" incorpore uma cláusula condicional para significar um imperativo hipotético. O fato de expressar ou não um imperativo deste tipo depende da razão que justifica a sentença no modo imperativo. Se a razão para alguém seguir a prescrição enunciada (i. e., uma razão para obedecer uma ordem) puder ser expressa por um enunciado condicional - que indica a base da obrigação de fazer o que o imperativo prescreve - do tipo "Fazer x (a ação prescrita) é um meio para você (a pessoa que quer seguir a prescrição) satisfazer um dos seus desejos", então aquele enunciado é um imperativo hipotético.

<sup>197</sup> *Idem.*

Portanto, segundo a teoria ética de Hobbes, a única razão para alguém *dever* fazer o que as "leis da natureza" prescrevem - para estar obrigado a obedecê-las - é o auto-interesse. Igualmente, a única razão para aceitar os imperativos morais como prescrições - (supondo uma "concepção internalista da racionalidade prática") para motivar alguém a obedecer, isto é, *querer fazer* o que os imperativos prescrevem - também é o auto-interesse. Pressupondo-se estas considerações, as "leis da natureza" são instanciações da seguinte forma geral:

*"Se se quer x, deve-se fazer y, onde x é a paz, um fim instrumental desejável na medida em que é necessário para a autopreservação, e y é a ação que, quando realizada em conjunto com a realização, por parte dos outros, da mesma ação, ajudará a produzir a paz".<sup>198</sup>*

A paz é um resultado (ou consequência) buscado por um tipo especial de combinação das ações realizadas por uma multiplicidade de indivíduos. A paz é um resultado *optimal*, quer dizer, as escolhas feitas pelos indivíduos são tais que, tomando o estado de natureza como situação inicial, quaisquer outras combinações que produzam mais ganhos para algum indivíduo, necessariamente produzem mais danos para algum outro indivíduo, e também é um resultado *estável* ou *em equilíbrio*, quer dizer, em relação ao estado de natureza, ninguém pode melhorar unilateralmente a sua situação, pois só pode ser individualmente mais vantajoso o resultado ligado a ações coordenadas (interdependentes).

Desta maneira, a paz não é um fim desejado em si mesmo, um fim que poderia ser dito "puramente racional", mas desejado porque serve como instrumento de garantia da autopreservação. Pode ser dito um fim ou bem comum apenas com a qualificação de que garante a cada um - dado o seu desejo natural<sup>199</sup> de autopreservação -, distributivamente, o bem (ou fim) mais desejado: a própria preservação. Assim, dada a *natureza humana* - i. e., a condição de indivíduos que buscam apropriar-se de todas as coisas que julgam capazes de

<sup>198</sup> Interpretação de Jean Hampton, baseada no texto do *Leviathan*. A consideração que Hobbes faz, no *De cive*, da chamada "lei fundamental da natureza", da qual "todas as outras são derivadas", pode ser considerada como oferecendo uma sustentação adicional àquela interpretação. Cf II, II, 92.

<sup>199</sup> "Natural" significa basicamente algo criado pela vontade de Deus - i. e., pelo Seu poder de causar todas as coisas - e que é introduzido como elemento da natureza, e isto quer dizer, como parte operante de algum *mecanismo* cuja manutenção não é de modo algum afetada por Deus; no caso, como elemento da natureza própria dos seres humanos.

maximizar a sua própria preservação, e cujo traço essencial da interação com cada um dos demais indivíduos é a liberdade [direito de natureza] de fazer, sem qualquer limitação do seu comportamento, tudo o que julgarem individualmente como necessário para atingir tal fim -,<sup>200</sup> um *bem comum* é algo que é racional para todo homem querer, quaisquer que sejam os seus desejos.<sup>201</sup>

### 5.6. A forma geral das "leis da natureza".

Se delimitamos o domínio das ações *y* àquelas que, quando realizadas multilateralmente, produzem a paz, então é possível apresentar a forma geral das leis da natureza de modo mais simplificado que o anteriormente apresentado:

Faça *y*, sempre que os outros também o fizerem (onde *y* pertence ao domínio das ações interpessoais que, quando realizadas coletivamente, produzem a paz; isto é, se satisfeito este critério, *y* pertence ao domínio das ações morais)<sup>202</sup>.

A segunda parte do enunciado indica uma condição necessária requerida para que a adesão do agente à lei que exige a realização de *y* lhe seja *obrigatória*. Por isso, as leis da natureza não são regras morais prescritas unilateral e incondicionalmente. Ou ainda, elas não podem ser derivadas - como mesmo Hobbes parece erroneamente sugerir, ainda

<sup>200</sup> Gauthier tenta expressar a natureza *não-social* das relações humanas que definem o conceito de *estado de natureza* (i. e., relações entre os homens *qua* indivíduos) no "contratualismo radical" de Hobbes do seguinte modo: "De fato, para Hobbes, as relações entre seres humanos são de dois tipos: relações de hostilidade, válidas no estado de natureza e constituintes deste estado; e relações de contrato, válidas no estado de sociedade e constituintes deste estado". *Moral dealing*, p. 328.

<sup>201</sup> A partir de uma passagem de *Rawls on justice*, de Thomas Nagel. Para uma outra análise do conceito de *bem comum* empregado por Hobbes, consultar Hampton, *op. cit.*, p. 46.

<sup>202</sup> Pressuponho que qualquer definição filosófica moderna de "*moralidade*" inclui o seguinte "traço mínimo": "a exigência recíproca de determinadas ações e omissões"; do que resulta a busca de *normas intersubjetivas* que as satisfaçam. Cf. Tugendhat, *Probleme der Ethik*, p. 67 na ed. espanhola. As filosofias morais antiga e moderna são construídas a partir de diferentes concepções sobre o que é, em geral, o objeto da ética: "a formulação da pergunta da ética antiga era: o que é que quero verdadeiramente para mim; a da ética moderna é: o que é que devo fazer com respeito aos outros". *Ibid.*, p. 51.

Ações morais são, num sentido, *ações coletivas*: dizem respeito ao tratamento que cada um, como membro de uma comunidade, deve dispensar aos demais, quer dizer, como deve "responder às necessidades e ambições dos demais". Ronald Dworkin, *Foundations of liberal equality*, p. 68 na ed. espanhola.

que sem usar o termo - da chamada "Regra de Ouro" (*Golden Rule*), regra que serviria de *teste* para saber se uma determinada lei da natureza aplica-se ou não em um dado caso particular [isto é, regra que enuncia as condições necessárias e suficientes para a (justificação racional da) adesão de alguém à lei da natureza que prescreve uma ação particular em questão]. No *De cive* e no *Leviathan*, tal regra aparece respectivamente nas suas formas negativa e afirmativa:

"... não se deve fazer aos outros o que não desejaríamos que fizessem a nós mesmos"<sup>203</sup>

"Faz aos outros o que gostarias que te fizessem a ti"<sup>204</sup>

Hobbes acrescenta um comentário sobre o conteúdo desta regra que pode aumentar a má compreensão do porquê é racional aderir às leis da natureza:

"O que ... cada um ... tem a fazer é apenas, quando ao comparar suas ações com as dos outros estas últimas parecem excessivamente pesadas, colocá-las no outro prato da balança, e no lugar delas as suas próprias, de maneira que suas próprias paixões e amor de si em nada modifiquem o peso"<sup>205</sup>

Se a racionalidade das leis da natureza estivesse ligada, como parece indicar Hobbes, à *universalizabilidade* ou *imparcialidade* destas leis, i. e., se a aceitação da validade das leis da natureza por um agente racional devesse pressupor necessariamente a adoção da perspectiva de outrem como sua *própria*, então não teríamos por que continuar a falar de *razões prudenciais*, pois desse modo a justificação de uma ação pessoal necessariamente deveria levar em conta a satisfação do interesse de outrem por meio da realização da ação, e isto excluiria a possibilidade de atribuir aos interesses de outrem um "peso" menor pela simples razão de pertencerem a outrem, e não à pessoa que realiza a ação.<sup>206</sup>

<sup>203</sup> III, XXVI, 117.

<sup>204</sup> 14, 64. No *De cive*, IV, XXIV, Hobbes lembra que a "Regra de Ouro" é "a lei de Cristo", e cita a passagem bíblica: *Mateus*, 7, 12. Nesta passagem, Jesus atribui o preceito a Moisés. Ver *Levítico*, 19, 34.

<sup>205</sup> *Leviathan*, 15, 79; grifo meu.

<sup>206</sup> Restaria ainda examinar o tipo de *imparcialidade* própria dos juízos morais da teoria ética hobbesiana. Tal como o vejo, este exame deveria começar pela consideração de que estes juízos não são *completamente universalizáveis*. Este último predicado poderia ser esclarecido a partir da apresentação da concepção de *universalizabilidade*, defendida por Richard Hare como uma propriedade essencial dos juízos morais: "*imparcialidade*", quando aplicado ao caso dos juízos

Existe, ancorando a “regra de ouro”, um pressuposto forte a respeito do acordo entre juízos de valor feitos por diferentes indivíduos. Pressupõe-se que os homens coincidem nos juízos sobre o que consideram como bom. Na sua forma negativa, essa regra proíbe fazer aos outros o que é indesejável a cada um. Mas, se alguém não gostaria, em certas circunstâncias, de ser lisonjeado, preferindo a franqueza ao engano, por que pressupor, por exemplo, que para a maioria dos homens uma agradável inverdade (supondo que, pelo menos algumas vezes, é disso que se trata a lisonja), nas mesmas circunstâncias, é indesejável?

Na sua forma positiva, a “regra de ouro” exige de cada um fazer aos outros o que considera desejável que, ao se conceber no lugar deles, fizessem a si mesmo. Aqui, a identificação de valores pode ter conseqüências bastante desvantajosas. Um exemplo extremo é apontado por Hans Kelsen: “se devemos tratar os outros como queremos ser tratados, fica excluída toda punição de um malfeitor, pois nenhum malfeitor deseja ser punido. Deste modo, é afastada parte essencial do direito positivo”.<sup>207</sup>

A regra que nos prescreve “tratar os outros do mesmo modo que desejaríamos que nos tratassem se estivéssemos no lugar deles” não requer dos outros que eles ajam do mesmo modo que nós agimos. Esta última exigência, a de *mútua adesão* a uma lei da natureza, é o que torna, com suficiência, para cada parte que adere à lei, *racional* a adesão mesma. Isto é, uma ação y só é racional quando os outros também a realizam porque, se realizada isoladamente, colocaria o agente à mercê dos outros que se omitissem de fazê-la, e isto seria contrário à própria função de uma lei da natureza, a saber, indicar uma ação que é eficaz na produção da paz e, *ipso facto*, na garantia do fim mais desejado: a autopreservação. Por isso, uma melhor interpretação da regra que pode ser considerada como *princípio fundamental* da teoria ética de Hobbes é a seguinte:

---

morais, significa um traço próprio de juízos completamente universalizáveis. Hobbes aponta - no *De cive* - uma oposição significativa entre “parcialidade” e “razão”. Devemos interpretá-la de forma mais restrita: é uma sugestão abreviada de que os *juízos prescritivos* da ética devem o seu traço de *imparcialidade (universalidade)* ao fato de serem feitos sob *condições de imparcialidade* (em algum sentido *apropriado* deste termo, i. e., *compatível com o critério* contido na regra de reconhecimento que representa “a forma geral das leis da natureza”).

<sup>207</sup> *O problema da justiça*, p. 19. Estas últimas observações sobre a “regra de ouro” inspiram-se na passagem que contém a citação de Kelsen.

"Faça aos outros o mesmo que eles fazem a você"<sup>208</sup>

Se atentarmos mais diretamente à discussão de Hobbes sobre a *natureza do dever ou obrigação moral*, poderemos nela encontrar uma base para a confirmação da interpretação de Kavka. No *De cive*, Hobbes - como parte da explicação para a ilusão dos homens sobre a racionalidade da violação unilateral das leis da natureza - indica com precisão que a base da aceitação de uma prescrição como obrigação moral é o fato de ela indicar uma ação que melhor promove o interesse de quem a ela adere:

"Pois toda infração da leis da natureza resulta do raciocínio falso, ou da tolice dos homens, que não se apercebem que os deveres e os serviços que prestam aos outros retornam a eles mesmos e são necessários a sua própria conservação"<sup>209</sup>

Nesta passagem, Hobbes indica qual é fonte do *poder motivacional* das leis da natureza: a *força* que elas têm de *obrigar* alguém a agir a partir do que elas estabelecem é derivada diretamente da eficácia das ações que prescrevem em promover do melhor modo possível os interesses de quem assim agir.<sup>210</sup> Porém, não posso deixar de assinalar uma característica importante que definem as circunstâncias sob as quais é racional obedecer às leis da natureza.

Um traço importante da interação humana é que, geralmente, "os deveres e os serviços" que cada um presta aos outros não retornam - para usar os termos de Hobbes - simultaneamente. Quer dizer, há quase sempre um intervalo não desprezível de tempo entre o momento em que se cumpre a própria parte num acordo e o aquele em que se recebe a contrapartida correspondente. É isso que torna, fato que não é possível de ser superado mesmo "fora" do estado de natureza, as expectativas razoáveis que se tem do cumprimento de acordos por parte dos outros um elemento essencial das condições de validade das leis

<sup>208</sup> Kavka, autor desta interpretação, ao comentá-la, faz ironicamente uma crítica certa à rejeição do tipo de justificação envolvido no uso de tal princípio. Essa justificação é criticada pelo simples fato de revelar a moralidade como constituída por comportamentos auto-interessados:

"Por que ela brilha menos vivamente enquanto ideal inspirador da conduta moral que o oferecido pela Regra de Ouro, eu a denomino de Regra de Cobre". *Right reason and natural law in Hobbes's ethics.*, p. 125.

<sup>209</sup> II, I, 91-92; grifo meu.

<sup>210</sup> Esta derivação deixa de ser estranha se lembramos que a teoria ética hobbesiana é *conseqüencialista*. Ver 1.1.



da natureza. Isto é o que Hobbes precisamente assinala numa passagem em geral ignorada pelos comentadores:

"(...) as leis da natureza não precisam ser publicadas nem proclamadas, pois estão contidas nesta única sentença, aprovada por todo o mundo: [Não façam aos outros o que não consideras razoável que seja feito por outrem a ti mesmo.]<sup>211</sup> alt.

Podemos agora enunciar mais precisamente a forma geral das leis da natureza por meio do seguinte *imperativo condicional*:

Se consideras razoável esperar reciprocidade na ação dos outros, ~~faça x.~~ faz x.

### 5.7. O conceito de racionalidade moral.

Há uma disputa importante para saber se uma obrigação justificada por razões prudenciais pode ser "genuinamente moral". Para Thomas Nagel, entre as razões capazes de justificar a adesão a regras morais (i. e., entre as notas do conceito de *racionalidade moral*) estão as chamadas *razões altruísticas*, razões para promover *diretamente* os interesses dos outros, sem considerá-los como meios para a promoção dos interesses *próprios*.

Peter Singer, um dos opositores da posição defendida por Nagel, rebate dizendo que "não há nada de irracional em estar interessado na qualidade da própria existência de um modo que não se está interessado na qualidade da existência de outros indivíduos. ... A racionalidade, no sentido amplo que inclui auto-consciência e reflexão sobre a natureza e valor da nossa própria existência, pode nos impulsionar em direção a interesses mais amplos que a qualidade da nossa própria existência; mas este processo não é necessário, e aqueles que não tomam parte nele - ou, ao tomar parte, não o seguem até o ponto de vista ético [i. e., o ponto de vista de um *espectador imparcial*] - não são irracionais nem mesmo estão errados".<sup>212</sup>

, 16<sup>a</sup>, na minha edição

<sup>211</sup> *Leviathan*, 26, 140. [Sublinhado meu.] Vale notar que aquilo que o agente julga como *razoável* alt. fazer depende das circunstâncias em que ele se encontra.

<sup>212</sup> *Practical ethics*, p.219.

Ao fazer menção a essa controvérsia, não tenho outra intenção senão a de adotar o partido de que não há nada de intrinsecamente errado com a concepção hobbesiana da racionalidade das ações guiadas pelas regras morais, as chamadas "leis da natureza". Assumo que a resposta que Hobbes oferece à pergunta "por que cada um deve agir moralmente?" é uma resposta plausível e não pode ser descartada pela simples (e *prima facie* pouco convincente) razão de que não estaria incluída entre as concepções "genuinamente morais" de racionalidade, concepções que estão de prévio acordo com o que pressupomos ser o "ponto de vista ético". Fazer isto equivaleria a descartar uma teoria não porque oferece uma explicação fraca para um tipo de fenômeno, mas simplesmente porque é uma explicação construída de modo distinto a todas as outras (que conhecemos) sobre os mesmos fenômenos.

Podemos recapitular pelo menos dois tipos de argumentos usados na tentativa de desqualificar as regras que Hobbes apresenta para elucidar a natureza da moralidade. O primeiro tipo de argumentação usado consiste em mostrar (às vezes sem apresentar boas razões para isso) que existe uma incoerência entre aqueles preceitos e as opiniões morais comuns (*common moral opinions*)<sup>213</sup> ou, para dizê-lo de modo ligeiramente distinto, mostrar que aqueles preceitos chocam-se com a moralidade do senso comum (*morality of Common Sense*).<sup>214</sup> O argumento pressupõe que duas funções básicas de uma teoria ética<sup>215</sup> qualquer são a *explicação* dos juízos morais comumente feitos - i. e., poder descrevê-los como aplicações a diferentes circunstâncias particulares dos princípios morais contidos na base da teoria - e a *justificação* desses mesmos juízos por meio de princípios, o que pode implicar a exclusão de alguns destes juízos como (normativamente) *inválidos*.<sup>216</sup>

<sup>213</sup> Expressão de Sidgwick, *op. cit.*, p. 423.

<sup>214</sup> *Idem*, p. 86.

<sup>215</sup> Como para qualquer outra teoria, a relação entre o que é prático (o que se prescreve) e o que é positivo (o que se descreve) é uma relação de implicação mútua. No caso da ética, os princípios podem "resumir" os fenômenos, mas estes, quando anômalos (quando colidindo de algum modo com o que é prescrito), podem implicar ou bem numa alteração dos princípios ou bem numa invalidação de algum juízo moral particular. Ver Sidgwick, *op. cit.*, principalmente pp. 1-2 e 373-374.

<sup>216</sup> Um exemplo deste tipo de estratégia pode ser encontrado em Richard Cumberland, moralista britânico que, ainda no séc. XVII (*On the laws of nature*, 1672), formulou uma das primeiras teorias éticas cujo método intuicionista era tomado como a principal arma contra a validação das teses de Hobbes. A pressuposição básica deste método pode ser assim resumida: os princípios éticos, que prescrevem o que é *dever* fazer incondicionalmente, uma vez enunciados de modo apropriado, expressam proposições auto-evidentes, isto é, proposições cuja verdade a *razão*, por meio de alguma operação, é capaz de apreender imediatamente, intuitivamente. Para Cumberland, existe somente um

O outro tipo de argumentação consiste na desqualificação direta dos preceitos da teoria hobbesiana, independentemente de serem falhos como base da explicação dos juízos morais do senso comum. Assim, os preceitos pretensamente morais da teoria hobbesiana não seriam, contudo, mais que meros "preceitos egoístas", quer dizer, formariam um sistema de regras cuja base é o auto-interesse. Thomas Nagel adota este tipo de argumentação:

*"Penso ter mostrado que, no sistema de obrigações desenvolvido no Leviathan, aquilo que Hobbes chama de obrigação moral está baseado inteiramente no auto-interesse ... Desta maneira, o Leviathan diz aos homens como podem agir para promover melhor os seus próprios interesses, de tal modo que se todos agirem da mesma forma, irão todos beneficiar-se ao mesmo tempo. Se Hobbes tivesse provado que algumas instituições e práticas promovem o bem-estar geral, e tivesse então dito que todo homem está obrigado a trabalhar pelo seu estabelecimento independentemente de elas promoverem ou não o seu interesse particular, então o conceito de obrigação aí envolvido poderia ser dito moral. Mas nada é dito neste sentido... Como uma tentativa de analisar conceitos morais, o Leviathan falha. Porém, caso se queira construir um sistema de comportamento político e social esperando que todos os homens considerem-no viável, parecerá natural apelar para o tipo de interesse racional de longo alcance sobre o qual Hobbes assenta o seu sistema."<sup>217</sup>*

---

princípio moral, que é o que prescreve a maximização do "bem comum de todos os seres racionais". De acordo com Sidgwick, ao sustentar que o 'bem comum de todos os seres racionais' é o *fim último* - entendido como o que é "Bom ou Desejável absolutamente e sem qualificação", isto é, como um "fim ... para o qual a ação de um agente racional como tal deve (*ought*) estar dirigida"; Sidgwick, *op. cit.*, p. 421 - cujo meio é a conduta guiada pelas regras morais, Cumberland pensava estar defendendo a moralidade do senso comum contra as "perigosas inovações de Hobbes". *Idem*, pp. 86 e 423. Estas inovações, contra as quais se voltava o utilitarismo de Cumberland, têm a sua base no próprio 'axioma do egoísmo racional': é racional (e correto) para uma pessoa, em quaisquer circunstâncias, agir de modo a produzir a maior felicidade própria. Contudo, tal estratégia de argumentação tem basicamente dois pontos fracos:

a) a dificuldade de determinar qual é afinal o mecanismo que opera na apreensão dessa verdade auto-evidente;

b) a dificuldade de explicar a possibilidade de os indivíduos motivarem-se a agir obedecendo ao princípio ético sempre que fazê-lo - dadas as circunstâncias que descrevem e limitam a situação de escolha - implica em contrariar o seu próprio interesse e inclinações naturais (interesse e inclinações pressupostos pela moralidade do senso comum, da qual a teoria parte e a qual pretende explicar). Ver Eduardo Rabossi, *Naturaleza humana y moralidad*, p. 126 e Sidgwick, *op. cit.*, pp. 96-103. O artigo de Rabossi traz uma exposição crítica mais detalhada do modelo de teoria ética adotado pelos chamados "racionalistas do séc. XVII" (dentre os quais, o autor inclui Cumberland) para "neutralizar" as teses éticas hobbesianas; pp. 119-126.

<sup>217</sup> *Hobbes's concept of obligation*, pp. 81-82.

Hobbes, no entanto, consideraria que "um sistema de comportamento político e social" (na verdade, de um sistema de regras prescrevendo condutas tipificadas) só poderia ser considerado viável por todos os homens quando pudermos supor que todos, ainda que divergindo nos seus juízos de valor (sobre o que é agradável e sobre o que é racional) particulares, compartilham o traço motivacional básico de agir obedecendo a uma regra somente quando fazê-lo for conducente à realização do seu auto-interesse. Esta coincidência de motivação torna possível que todos os homens compartilhem pelo menos um juízo de valor: "a paz é uma boa coisa".<sup>218</sup> A justificação - capaz de ser encontrada por cada homem, quando capaz de determinar o melhor meio para alcançar o que mais deseja - para este acordo é a seguinte: a paz é um meio para a obtenção da própria preservação (concebida como uma maximização do tempo de vida e da satisfação dos desejos durante ele formados).

#### 5.8. O que há de errado com o egoísmo?

Mas há uma maneira mais clara de mostrar se existe e onde uma inconsistência no egoísmo. Alguns autores tentaram encontrar pontos de auto-contradição na concepção egoísta. George E. Moore disse que esta concepção, "que sustenta que cada um de nós deve perseguir a nossa maior felicidade como nosso fim último", é uma "flagrante contradição".<sup>219</sup> Moore baseia sua crítica na tese de que há uma confusão, ainda que não aparente, na noção de "meu próprio bem", pois é um absurdo pressupor que existe algo que pertence *exclusivamente a mim*, um *bem privado*, tanto quanto o é dizer que algo pode *existir* privadamente, ou somente *para* uma pessoa. O argumento de Moore é resumidamente este:

(a) para dizer significativamente que "*A* é meu próprio bem" devo pressupor que "existe um *A* tal que *A* é bom e *A* pode ser meu, pertencer a mim";

(b) "*A* é bom" significa "*A* é absolutamente bom";

(c) "Que *A* pertença a mim é um fato bom" ou "*A* é o meu próprio bem" pressupõem que "*A* é absolutamente bom";

<sup>218</sup> *Leviathan*, 15, 80.

<sup>219</sup> *Principia ethica*, pp. 96 e 102; citado por Gauthier, *Moral dealing*, p. 234.

(d) A única razão que tenho para dever buscar *meu próprio bem* é o fato de *A* ser *absolutamente bom* (ou, indiretamente, que é *absolutamente bom* o fato de possuir *A*);

(e) Se *A* é absolutamente bom, então as demais pessoas tem igualmente tanta razão quanto eu para dever buscar o *meu próprio bem*;

(f) Logo, é falso que o fim último da conduta racional de cada um é a sua *própria* felicidade.

Este suposto argumento *ad hominem* só funciona, no entanto, se o *egoísta* - ao afirmar que "é racional para todo indivíduo buscar somente a própria felicidade" ou que "a felicidade de cada indivíduo é o único bem para ele" - também estiver assumindo que "sua própria felicidade" (a satisfação de seus próprios desejos) é algo *absolutamente* (ou objetivamente) bom. No entanto, não é necessário que o *egoísta* faça esta pressuposição. Não é necessário que a faça simplesmente porque ele parte da concepção de que o predicado bom é sempre relativo, quer dizer, é sempre relacional: "... é bom para ...". Se um *egoísta* aceitar o passo argumentativo dado em *b*, sua doutrina será fatalmente refutada. Mackie comenta o ponto fraco do argumento de Moore:

*"... Moore está simplesmente asserindo que qualquer coisa boa que possa constituir uma razão para qualquer pessoa buscá-la deve ser 'absoluta' - quer dizer, que 'bom' deve ser um predicado monádico. Contudo, ele não tem nada para sustentar esta asserção exceto suas opiniões extraordinariamente limitadas sobre as formas possíveis das construções lingüísticas significativas"*<sup>220</sup>

Thomas Nagel também reconheceu a mesma deficiência naquele argumento:

*"[As afirmações de Moore] parecem-lhe auto-evidentes porque ele considera como já assentado que 'bom' é um predicado monádico que denota uma propriedade não-natural simples"*<sup>221</sup>

Henry Sidgwick mostrou pela primeira vez por que qualquer tentativa de refutação do egoísmo por meio de argumentos do tipo *ad hominem*- mostrando essencialmente a existência de alguma contradição interna à teoria *egoísta* - está fadada ao fracasso. A razão é a de que qualquer argumento *ad hominem* cogente contra o egoísmo só

<sup>220</sup> Sidgwick's pessimism, p. 322.

<sup>221</sup> The possibility of altruism, p. 86, nota 2.

pode ser construído se o egoísta introduzir na sua doutrina a noção de algo que é absolutamente bom. Ora, o egoísta não tem necessidade de incluir tal noção em sua doutrina.

Seguindo a posição adotada por Gauthier, creio que é por meio do uso da estrutura analítica oferecida pela teoria da escolha racional - ou mais precisamente, pela parte dela que é denominada de "teoria dos jogos", isto é, "uma teoria da interação racional entre dois ou mais indivíduos, cada um dos quais perseguindo racionalmente seus próprios objetivos contra o(s) outro(s) indivíduo(s) que persegue(m) racionalmente os seus próprios objetivos"<sup>222</sup> - que se pode fazer aparecer algo de errado na "concepção 'de maximização da utilidade' da racionalidade prática".<sup>223</sup> Gauthier resume assim o seu projeto:

*"Como um filósofo, tenho uma carta na manga que, senão é realmente paradoxal, deve parecer inesperadamente intrigante. Porém, as questões que levantarei sobre o egoísmo vêm, não do repertório filosófico tradicional, mas antes da teoria da escolha racional. Mais particularmente, ainda que o egoísta solitário venha a ser aprovado no escrutínio racional, quando situado com outros da mesma crença, numa interação na qual cada um busca maximizar a sua própria felicidade, surgem razões para contestar a racionalidade do egoísmo. E estas razões dizem tanto respeito à preocupação do egoísta com a sua própria felicidade, como o dizem ao princípio maximizador da escolha. Há algo errado com a nossa concepção de racionalidade prática"<sup>224</sup>*

<sup>222</sup> John Harsanyi, *Morality and the theory of rational behaviour*, in Amartya Sen & Bernard Williams, *Utilitarianism and beyond*, p.43. A teoria dos jogos caracteriza-se mais especificamente por "cuidar de problemas onde o ganho de cada participante não depende apenas das suas decisões e dos estados da natureza, mas também das ações de outros participantes inteligentes". Mário Henrique Simonsen, *Ensaio analítico*, p. 407.

<sup>223</sup> Ver 4.10.

<sup>224</sup> *Moral dealing*, p. 235.

### 5.9. O projeto de reconciliação da prudência com a moralidade.

Voltando ao exame interno da teoria ética hobbesiana, torna-se claro que ainda não avaliamos um problema mais radical: saber se Hobbes foi bem-sucedido no seu projeto de reconciliação da prudência com a moralidade. Este projeto é desenvolvido visando apresentar uma resposta possível para um problema mais geral, um problema nuclear da filosofia moral:

*"Desde o tempo do desafio de Glaucon a Sócrates, os filósofos morais tentaram mostrar que agir moralmente é agir em nosso auto-interesse racional. Eles tentaram mostrar que, ao perseguirmos nosso auto-interesse, compreendido corretamente, estaremos comportando-nos moralmente, e que, ao agirmos moralmente, nosso auto-interesse racional estará sendo melhor atendido. Eles tentaram mostrar que a pessoa moral será a mais feliz"<sup>225</sup>*

A solução hobbesiana pressupõe que compreender corretamente o que é racional para o ser humano querer para si (i. e., se interpretamos corretamente, usando uma teoria sobre a natureza humana, em que consiste o "nosso auto-interesse racional") significa compreender que o fim último da conduta racional de cada ser humano é a sua própria preservação,<sup>226</sup> mesmo que, como indivíduos, não concordem sobre o modo mais valioso de viver.

Fazendo tal pressuposição, para se poder apreciar adequadamente a solução hobbesiana, é preciso investigar que condições devem ocorrer para que possamos esperar com certeza, i. e., seguindo o plano das provas oferecidas pela teoria ética, a adesão de cada agente que se guia por cálculos prudenciais racionais - um agente que calcula o mais precisamente possível o que deve fazer para produzir *para si* as conseqüências mais benéficas - a regras que são eficazes somente quando seguidas coletivamente, i. e., quando seguidas também por outros indivíduos.<sup>227</sup>

<sup>225</sup> Robert Nozick, *Philosophical explanations*, p. 405.

<sup>226</sup> A autopreservação é o bem individual, se definimos "bem" como "aquilo que, de acordo com a razão, é fundamental visar". Mas tal definição deve ser bem interpretada para o caso do "egoísmo ético". Segundo dois pensadores modernos pelo menos - Spinoza e Hobbes - a proposição "a autopreservação é o único fim último da conduta racional de cada indivíduo" é uma tautologia, porque o "egoísmo ético" assume que todo indivíduo racional (i. e., todo ser humano que persegue racionalmente seus objetivos) busca a maior felicidade própria como fim último da sua conduta. Ver Sidgwick, *op. cit.*, pp. 89-93.

<sup>227</sup> Não parece ser necessário expressar esta última condição como "todos os demais indivíduos".

*nos diz nte.  
resta a base  
C/Hobbes se  
um caso de  
avacouder*

Uma concepção contemporânea sobre a natureza das regras morais e da moralidade - entendida aqui como regularidades de comportamento resultante da adesão coletiva àquelas regras - que repete um traço central da concepção hobbesiana é adotada por Kurt Baier: "a verdadeira *raison d'être* de uma moralidade é produzir razões superiores às razões auto-interessadas nos casos em que seria prejudicial a todos se todos seguissem o seu auto-interesse".<sup>228</sup>

Podemos dizer que as regras morais são "coletivamente racionais", quer dizer, todos podem esperar melhorar a sua situação no caso de todos cumprirem o que elas ordenam, comparativamente, pelo menos, às suas perspectivas na ausência de qualquer acordo, isto é, na permanência da situação coletiva descrita como "estado de natureza".<sup>229</sup>

O projeto de reconciliação pode ser avaliado com mais clareza a partir de uma elucidação da concepção de Hobbes sobre a *validade* das "leis da natureza". Começo enunciando as *condições de validade* dos princípios éticos da teoria de Hobbes: as "leis da natureza". Elas podem ser divididas em três grupos:

a) as condições que expressam as razões que justificam os princípios como os melhores meios para cada agente alcançar a paz (digamos, as "*condições de justificação*" dos princípios, que explicam por que é racional segui-los);

b) as condições subjetivas graças às quais os agentes são capazes de aplicar os princípios normativos a eles mesmos ao formarem os seus *motivos para agir* (digamos, as "*condições de motivação*" dos agentes, que explicam por que os princípios possuem força motivacional sobre um agente racional - i. e., para o qual é racional seguir tais princípios);

c) as condições objetivas sob as quais a *obediência* aos princípios está justificada (digamos, as "*condições de eficácia*" dos princípios, que indicam as circunstâncias que garantem que a paz - o fim desejado pelos agentes racionais que aderem voluntariamente aos princípios - seja o resultado de fato obtido).

As condições referidas em *b* incluem capacidades subjetivas (Hobbes diria, "interiores") dos agentes racionais:

b.1. a capacidade de chegar a conhecer os princípios através de *retos raciocínios* sobre o que deve ser feito para produzir a paz;

b.2. a capacidade de incluir aquelas "conclusões" ou "teoremas da razão" como *princípios normativos* que ingressam na cadeia de motivações - uma deliberação - que leva à decisão de fazer o que é por eles prescrito.

<sup>228</sup> *The moral point of view*, p. 309; citado por David Gauthier, *Morals by agreement*, p. 10.

<sup>229</sup> Ver John Rawls, *A theory of justice*, pp. 269 e 496-497.



As referidas em *c* incluem as condições que tornam a adesão de cada agente aos princípios um comportamento capaz de produzir para si o melhor resultado esperado, i. e., a paz, estado no qual cada um pode razoavelmente esperar otimizar o que considera como mais vantajoso para si.<sup>230</sup> Estas condições incluem, portanto, a expectativa de cada indivíduo - que sabe estar sujeito (e supõe corretamente que os demais também o sabem) ao poder soberano - quanto à obediência dos demais às "leis da natureza". Que (pelo menos) a maioria das pessoas conforme as suas ações às prescrições contidas em tais leis é a condição *sine qua non* da paz; a condição de eficácia de toda lei da natureza. Dito de forma genérica, desde o ponto de vista da teoria, só podemos ter certeza da adesão de cada indivíduo às "leis da natureza" se tal indivíduo souber que pode esperar racionalmente que o comportamento dos outros indivíduos será guiado pelas mesmas leis.

As condições referidas em *b* e *c* são aquelas que garantem a estabilidade motivacional dos princípios; são condições necessárias para a produção da paz, o fim desejado por todo agente racional quando age guiado pelas "leis da natureza". As considerações em *c* não são totalmente objetivas, pois incluem a expectativa de cada indivíduo - se ele sabe estar sujeito (e supõe corretamente que os demais também o sabem) ao poder coercitivo de um soberano - quanto à obediência dos demais aos princípios éticos. É uma condição *sine qua non* da paz que (pelo menos) a maioria das pessoas conforme as suas ações às prescrições contidas naqueles princípios. Dito de forma genérica, só podemos ter certeza da adesão de cada indivíduo a tais regras (que são *convenções* restringindo o comportamento de "egoístas crassos")<sup>231</sup> se ele souber que pode esperar racionalmente que o comportamento dos outros indivíduos resultará da adesão às mesmas regras.

Deste modo, as leis da natureza são convenções eficazes (somente quando obedecidas coletivamente). As condições de possibilidade de uma *adesão coletiva* - isto é, o

<sup>230</sup> Quer dizer, o melhor resultado possível se medido pela escala de *utilidade individual*.

<sup>231</sup> Adotando uma terminologia consagrada a partir da obra *Convention*, de David Lewis, as "leis da natureza" podem ser ditas regras de conduta *convencionais* (ou simplesmente "*convenções*") no sentido de que o comportamento por elas orientado é em parte justificado pelo fato de cada agente saber que a maioria das pessoas igualmente se orienta assim, e também saber que esta maioria espera, com boas razões para isso, que a maioria das demais pessoas aja conformando suas ações às mesmas regras. Ver Gauthier, *Moral dealing*, p. 47. "Egoístas crassos" são os agentes que fazem o que pensam ser suficiente para maximizar seu benefício próprio, não aceitando nenhuma limitação, prescrita por alguma regra, neste comportamento.

fato de a maioria das pessoas comportar-se de acordo com as restrições prescritas pelas leis da natureza - são as seguintes:

a) cada pessoa *prefere* (i. e., dada a sua psicologia e interesses, deve ter uma melhor razão para *preferir*) que (pelo menos) a maioria das pessoas adira às leis da natureza a que as ignore (isto define a prática - i. e., a regularidade *R* de comportamento - interativa dominada por adesões mútuas às leis da natureza como uma "prática beneficentemente coordenativa"; sendo a adesão mútua a elas uma condição necessária para os homens "escaparem" do estado de natureza)<sup>232</sup> e também *prefere* não aderir a essas leis nos casos em que a maioria das outras pessoas também não o faça (isto define a prática interativa dominada por adesões mútuas às leis da natureza como uma "prática coordenativa não-dominante", pois a sua preferência por aderir às leis não representa a sua melhor resposta independentemente do que os demais fizerem);

b) dado o conhecimento comum de que (pelo menos) a maioria das pessoas conforma as suas ações às leis da natureza e tem boas razões para esperar que os outros também façam o mesmo, cada pessoa *prefere* (e, para isso, deve ter uma razão superior - relativamente a sua psicologia e seus interesses - para *preferir*) aderir às leis a ignorá-las.

As condições *a* e *b* indicam que as leis da natureza (a definição valeria para qualquer conjunto de *convenções*) são racionalmente estáveis, e isto quer dizer, satisfeitas as condições de validade antes enunciadas, cada agente tem razão para preferir aderir a tais regras. Se também for o caso que usualmente os agentes inclinam-se ao comportamento que expressa uma tal adesão, então as "leis da natureza" também são motivacionalmente estáveis.<sup>233</sup> É bom lembrar que a estabilidade motivacional destas regras depende da força normativa que elas possam ter relativamente a outros tipos de motivação, quer dizer, de disposições que de algum modo inclinam o agente a deixar de fazer o que tais regras prescrevem.

esta estabilidade da qual  
11 motivação

<sup>232</sup> Os benefícios apontados genericamente por Hobbes são: poder dispor das "coisas que são necessárias para uma vida [pacífica, sociável e] confortável ... [e poder] consegui-las através do trabalho". *Leviathan*, 13, 63 e 15, 80.

<sup>233</sup> Cf. distinções conceituais feitas por David Gauthier in *Moral dealing*, pp. 16-17 e 48-50 e *Morals by agreement*, pp. 11-12. Na verdade, estas afirmações sobre a preferência que define a *estabilidade racional* das leis da natureza escondem um problema importante. Preferências dependem de desejos de 2<sup>a</sup> ordem, e servem para orientar a escolha relativa a quais desejos buscar satisfazer primeiro; por isso a racionalidade de uma preferência está limitada pela hierarquia de desejos que ela estabelece. Quais são então, no caso da preferência para aderir às leis da natureza, os desejos que a tornam racional? Ou, para que tipo de agente - e sob quais circunstâncias relativamente a outros agentes - as leis da natureza são racionalmente estáveis? Espero que um esboço de resposta a tais indagações ganhe forma mais precisa deste ponto em diante desta dissertação.

### 5.10. Um acordo moral entre egoístas pode ser justificado?

Entre os comentadores interessados em minar o argumento hobbesiano justificando um acordo *moral* entre egoístas, há dois grupos que representam as posições mais radicais (i. e., que ocupam os extremos de uma série de variações possíveis). Um primeiro grupo inclui diferentes argumentos para mostrar que, dadas as premissas do modelo hobbesiano de estado de natureza, tal acordo é impossível. Um segundo, argumentos mostrando que se trata de um acordo desnecessário; quer dizer, que não é necessário postular um *estado de natureza* com as características hobbesianas como base da explicação sobre a natureza e necessidade de um contrato moral garantindo a existência, entre os homens, de uma rede mínima de interações cooperativas que abreviamos genericamente por *sociedade*.

Vou primeiro considerar exemplos pertencentes a cada uma dessas posições extremas e mostrar por que, pelo menos tais exemplos, não refutam o argumento hobbesiano. Posteriormente, restará um argumento mais refinado contra Hobbes, a saber, o de que a solução para o "problema do egoísmo" - expresso por um padrão de interação interpessoal que torna impossível a consecução de fins egoístas - não pode ser uma solução egoísta. Tal como assinalou Leslie Mulholland de forma precisa, para se poder desfazer o chamado "dilema do egoísmo" - quer dizer, a incapacidade da teoria egoísta de prover regras racionais para limitar os apetites de cada indivíduo na interação com os demais -, é necessário pressupor algo a mais que as meras premissas desta teoria. Veremos adiante se esta objeção é válida ou não relativamente à teoria hobbesiana.

Um exemplo de argumento pertencente ao primeiro grupo é o seguinte. No livro *Vícios privados, benefícios públicos?*, o economista Eduardo Giannetti da Fonseca afirma que o "grau de adesão interna dos indivíduos às normas de conduta que sustentam a vida comunitária" é zero no modelo hobbesiano. A seguir, diz que o grau de adesão não pode ser nulo "porque sem um mínimo de adesão, isto é, sem um acordo moral básico, construído de baixo para cima e razoavelmente internalizado por um grande número de membros da comunidade, a ordem social não pode existir".<sup>234</sup> Creio - como ficará mais claro pelo que se segue - que este autor está incorrendo numa confusão conceitual, a saber,

---

<sup>234</sup> P. 87.

entre as condições de estabilidade motivacional das "normas de conduta que sustentam a vida comunitária" e a racionalidade de tais normas. Esta racionalidade significa, naquele modelo, que a razão para alguém fazer o que tais normas prescrevem (não fazendo aqui diferença se tomamos "normas" por "leis da natureza" ou "leis civis") pode, em última análise, sempre ser expressa pelo condicional "Fazer o que tais normas prescrevem é o melhor meio para satisfazer o fim que mais se deseja, a *própria* preservação".

Se por "adesão interna às normas" este economista quer se referir à motivação para obedecer regras de conduta, não pode excluir o medo da punição como uma forma de explicar uma tal adesão; e também não pode falar, neste caso, de grau de adesão nulo. E, basta que o "preço" a ser pago pela violação das normas seja suficientemente alto para que seja racional o seu cumprimento (por parte de um indivíduo que calcula o que fazer para obter para si os maiores benefícios possíveis).

Respondendo a uma pergunta feita pelo próprio autor, "o que nos leva a acatar uma norma de conduta" é, pelo menos segundo Hobbes, o benefício que a conduta individual assim regrada, quando houver uma adesão coletiva à norma, proporcionará a cada indivíduo. Um "acordo moral básico" sobre um conjunto de regras de conduta é, em parte, um acordo sobre a racionalidade da adesão individual a estas regras, quer dizer, um acordo sobre as condições nas quais é racional para cada um seguir tais regras. Dizer, no caso do modelo hobbesiano, que "o acordo moral é imposto de fora e de cima para baixo pelo soberano"<sup>235</sup> significa dizer que as regras acordadas não são nem mesmo racionalmente justificadas para cada um dos acordantes, independentemente de ser ou não o caso que a maioria deles conforma as suas ações a tais regras. Sem esta justificação, não há diferença entre o tipo de "acordo moral" a que se refere este economista e uma imposição de normas de conduta pela violência ou força. Concluindo, um acordo moral deste tipo não é, no sentido próprio da palavra, um acordo *moral*, pois não pode haver obrigação moral "fora" do livre consentimento. Ou melhor - e isto é o núcleo do conceito de *autonomia*, presente em qualquer teoria ética moderna - *só se está obrigado ao que se quis livremente obrigar*. Deste modo, a legitimidade política tem só uma única fonte: a

<sup>235</sup> *Ibid.*, p. 86.

Individual  
fe  
sua  
de um  
total  
N  
qu  
da  
co  
de  
do  
de  
votante

livre declaração da vontade - por cada um dos contratantes - de limitar a sua própria liberdade.<sup>236</sup>

Pode-se considerar um traço das teorias medievais da "lei natural", pelo menos a partir de São Tomás de Aquino, a visão de que os seres humanos - a exemplo de todas as coisas no universo - têm um papel a cumprir numa ordem objetiva; tal papel determina, em última instância, os objetivos que cada indivíduo deve (porque são os *verdadeiros* ou *reais*) buscar atingir. Coerentemente com tal visão, a "moralidade era [considerada como] uma formulação da lei eterna de Deus: o ordenamento harmonioso das causas finais na natureza no que diz respeito especificamente aos seres humanos".<sup>237</sup> Em oposição, o ponto de onde partem as teorias *modernas* da "lei natural" é a "afirmação de que os indivíduos têm direito de determinar os seus próprios objetivos",<sup>238</sup> os seus fins particulares.

Em Hobbes, mais especificamente, a "força normativa" das leis da natureza ou leis morais - o poder que têm de obrigar - tem origem na própria natureza humana: apenas para seres que buscam racionalmente a *própria* preservação e que se reconhecem numa situação com certas características "contingentemente inescapáveis",<sup>239</sup> tais normas de conduta têm força motivacional, força associada ao motivo conclusivo ("último apetite") de uma deliberação. A "força normativa" das leis da natureza é explicada, em última análise, pela crença de que certas ações, sempre que guiadas coletivamente por elas, produzem condições sob as quais cada um pode melhor perseguir os seus próprios objetivos. E mesmo que estas leis possam não se ajustar totalmente a alguma concepção *intuitiva* (aqui significando, do *senso comum*) que tenhamos sobre a moralidade, é inegável que "a moralidade hobbesiana pode [ainda] ser o melhor que podemos esperar num mundo sem deveres naturais ou valores objetivos".<sup>240</sup>

Quero agora apresentar um exemplo de argumento pertencente ao segundo grupo. James Sterba, no artigo *Justifying morality: the right and the wrong ways*, atribui a Hobbes um modelo de "estado de natureza" que na verdade descreve a condição da

<sup>236</sup> Como assinalou Rousseau, a expressão "o direito do mais forte" não designa um direito, pois é uma contradição em termos. Por quê? "La force est une puissance physique; je ne vois point quelle moralité peut résulter des ses effets. Céder à la force est un acte de nécessité, non de volonté (...) En que sens pourra-ce être un devoir?" *Du contrat social*, I, III; p. 22.

<sup>237</sup> Stephen Darwall, *Motive and obligation in Hume's ethics*, pp. 415-416.

<sup>238</sup> J. B. Schneewind, *Modern moral philosophy in Singer* (ed.), *A companion to ethics*, p. 148.

<sup>239</sup> A expressão é usada por Stephen Darwall, *op. cit.*, p. 442, nota 3.

<sup>240</sup> Will Kymlicka, *The social contract tradition in Singer*, *op. cit.*, p. 191.

humanidade já "fora do estado de natureza", a *sociedade*. O argumento de Sterba parte de uma clarificação conceitual. Resumindo a sua exposição, deve-se começar por incluir no conceito de *egoísmo racional*, se queremos interpretá-lo de modo plausível, as seguintes teses:

(a) a moralidade - i. e., um sistema de regras que é racional para cada pessoa seguir (assumindo que o verdadeiro propósito de agir seguindo os ditames da razão seja o "de maximizar a satisfação e minimizar a frustração")<sup>241</sup> na interação com as outras - justifica-se pelo auto-interesse;

(b) "auto-interesse" é um modo abreviado de dizer "auto-interesse de longo prazo";

(c) um sistema de regras que é racional para cada pessoa seguir é o que indica um conjunto de ações cujo fim supremo, buscado deliberada e racionalmente, é a maximização da satisfação do auto-interesse de longo prazo;

(d) agir moralmente é o mesmo que agir "conforme o dever", i. e., conforme as regras que definem a moralidade; para isso, é suficiente estar motivado a adotar os melhores meios para realizar o auto-interesse, e não é necessário estar motivado pelo assim chamado "fim da moralidade", algo como querer agir "pelo dever" (seja porque isto é propriamente um ditame da razão prática, seja porque a razão para agir moralmente só existe para aqueles que visam atender diretamente aos interesses, intenções ou desejos dos outros - por considerarem a promoção do bem-estar dos outros como um valor intrínseco).

A seguir, Sterba sustenta - também seguindo Baier - que o egoísmo racional preenche uma importante exigência da razão prática: para todo  $y$ , é racional para  $y$  buscar a felicidade de  $y$  (a maximização da satisfação de  $y$ )<sup>242</sup>. Dadas estas pressuposições, apresenta uma conclusão apenas em parte acertada, já que não pode ser aplicada ao modelo hobbesiano de estado de natureza:

*" (...) os efeitos sobre a satisfação da vida do egoísta resultantes do fato de todos seguirem as diretivas do egoísmo racional podem não ser assim tão desastrosos quanto o foram algumas vezes considerados ... Um mundo onde todos buscam o auto-interesse de longo prazo sob arranjos institucionais ideais pode não ser um mundo tão ruim para se viver, ainda que fosse diferente, tal como argumentamos, de um mundo onde todos agissem pelo dever ou conforme ele. Portanto, a guerra de todos contra todos de Hobbes pode ser um mau modelo de um estado de coisas onde reina o egoísmo racional"*<sup>243</sup>

<sup>241</sup> Suposição assim apresentada por Kurt Baier; citado por Sterba, *op. cit.*, p. 50.

<sup>242</sup> Ver também 4.9.

<sup>243</sup> *Op. cit.*, p. 53.

Contudo, um mundo onde *a maioria* - não necessariamente *todos* - busca o auto-interesse de longo prazo (já que é possível que algumas vezes alguns ajam impulsivamente e violem as regras que não são do seu *próprio* interesse de longo-prazo violar) - um mundo onde reina, ainda que não sem exceções, o egoísmo racional (no sentido definido por Sterba) - não é o estado de guerra de todos contra todos, mas justamente o mundo resultante do acordo entre egoístas racionais para criar instituições sob cujos arranjos as regras do egoísmo racional são eficazes para a maximização do que cada um mais deseja, a felicidade *própria*.

### 5.11. A racionalidade coletiva da "disposição para a paz".

Para se construir uma teoria completa da razão prática, deve ser possível mostrar que as ações individuais estão de acordo com algum sistema de regras de comportamento que é aceito por um agente racional. Quer dizer, por todo e qualquer agente racional, tal como essa teoria o concebe. Isto implica que um critério necessário de racionalidade das ações individuais - no caso de uma teoria ética, de sua *correção* moral - é a coordenação das ações individuais. Assim, um critério negativo para a atribuição do predicado *racional* a ações individuais é: não se deve realizar uma ação cuja realização pela maioria não é coletivamente racional. A única "racionalidade coletiva" admitida como normativa pelo *egoísmo racional* é a baseada em interesses pessoais: ninguém deve realizar uma ação que, se também levada a cabo coletivamente, conduz a danos para si mesmo. No caso disso ocorrer na situação descrita por Hobbes como "o estado de guerra de todos contra todos", o mau resultado individual manifesta-se do modo mais insidioso: é pela generalização da violência, cuja conexão com a violência individual nem sempre é percebida,<sup>244</sup> que a natureza do próprio indivíduo passa a ser constantemente ameaçada.<sup>245</sup>

A única racionalidade coletiva é a baseada em interesses pessoais.

<sup>244</sup> Há, conforme já foi indicado, vários modos não excludentes de explicar esta deficiência de percepção: déficits de informação, por exemplo, o desconhecimento das conseqüências mais longínquas de ações violentas individuais; irracionalidade epistemológica, por exemplo, juízos baseados numa má compreensão da estrutura da situação enfrentada - o que pode envolver preconceitos, entendidos aqui como pressuposições sem boas evidências, sobre a inteligência e motivação dos outros agentes; predominância de crenças ligadas puramente a paixões, por exemplo,

Esta exigência faz com que parte essencial de uma teoria ética seja a explicação das "condições de motivação"<sup>246</sup> de um agente que aceita determinadas regras como racionais e guia sua conduta por elas. Nesta explicação são indicadas as motivações ou disposições necessárias e suficientes para se agir seguindo as regras aceitas como racionais, considerando que, em nenhuma circunstância, a adesão coletiva a tais regras pode ser coletivamente irracional.<sup>247</sup>

A exigência da "racionalidade coletiva" tem um fundamento metafísico forte - presente de modo dominante na tradição do pensamento ético e político inaugurada por Platão e que se aplica, entre outros, aos racionalistas do século XVII. Parte deste "ideal platônico" foi assim resumida por Isaiah Berlin: "as respostas verdadeiras, quando encontradas, devem necessariamente ser compatíveis entre si e formar um todo único, pois uma verdade não pode ser incompatível com outra - coisa que sabemos *a priori*".<sup>248</sup>

Como reconciliar, no caso da moral, os conflitos entre planos de vida racionais individuais? Numa sociedade capaz de evitar tais conflitos, não é de supor-se que "os homens racionais respeitarão o princípio da razão em cada outro homem, e não sentirão o desejo de lutar entre si ou de dominarem-se uns aos outros"?<sup>249</sup> O problema da "racionalidade coletiva" só pode ser solucionado, segundo a ética hobbesiana, por uma reorganização racional da sociedade que, sem necessariamente contar com a possibilidade de eliminar completamente o conflito de valores, garanta uma situação de compromisso mínimo sem o qual os seus membros mal sobrevivem, pois permanecem presas de extremos desespero, insegurança e sofrimento.

Na teoria hobbesiana, a "disposição para a paz" é o que faz com que uma pessoa tenha uma forte razão para preferir que (pelo menos) a maioria das pessoas adira às leis da natureza a que as ignore. Só deseja que as leis da natureza sejam cumpridas aquela pessoa que deseja a paz e crê que ela pode ser obtida se (pelo menos) a maioria das pessoas conforma as suas ações às leis da natureza. E, o que é mais importante, é esta "disposição

---

as considerações que não passam de *wishful thinking*; ou inconsistência de valores; por exemplo, desejar alcançar fins por meios que lhes são incompatíveis.

<sup>245</sup> Ver, por exemplo, *Leviathan*, 15, 73.

<sup>246</sup> Ver 1.1 e 5.9.

<sup>247</sup> Ver Julian Nida-Rümelin, *Practical reason, collective rationality and contractarianism*, in Gauthier, D. e Sugden, R. (eds.), *Rationality, justice and the social contract*, pp. 54-58.

<sup>248</sup> *Limites da utopia*, p. 17.

<sup>249</sup> Berlin, *Quatro ensaios sobre a liberdade*, p. 152.



para a paz" que explica por que as leis da natureza - nas palavras de Hobbes - "obrigam *in foro interno*".<sup>250</sup>

As leis da natureza são imperativos hipotéticos que informam o que um agente deve fazer, *se* deseja produzir a paz. Todavia, a paz não pode ser produzida por sua ação individual, mas sim por uma ação coletiva. Sempre que souber que há "garantia suficiente de que os outros observarão para com ele as mesmas leis" e, assim informado, deliberar guiado pelo desejo de produzir a paz, desejará por em prática as leis da natureza.<sup>251</sup> O que significa que estas leis nem sempre obrigam *in foro externo*, quer dizer, nem sempre "[impõem] o desejo de pô-las em prática".<sup>252</sup>

<sup>250</sup> *Leviathan*, 15, 79.

<sup>251</sup> Sob estas condições, "desejar a paz" é uma disposição da vontade descritível como uma "virtude", pois tal disposição consiste - para usar uma expressão cartesiana - na "firme e constante resolução de ... nunca carecer de vontade para empreender e executar todas as coisas que [se] julgue serem as melhores". *As paixões da alma*, terceira parte, art. 153.

<sup>252</sup> *Idem*. Por esta razão, as leis da natureza são apenas *ditames da razão* e não propriamente leis. Uma lei é uma *ordem* ou *ato de comando* de quem tem direito e poder de mando sobre outros. Cf. *Leviathan*, 15, 80. Em *A dialogue between a philosopher and a student of the common laws of England*, Hobbes afirma: "Por leis entendo leis vivas e armadas... Não é, portanto, a palavra da lei, mas o poder de quem tem a força de uma nação, o que faz efetivas as leis"; p. 10 na trad. espanhola. Há duas passagens no *Leviathan* que deixam mais clara a relação entre "lei" da natureza e a lei civil. Cito-as integralmente: "A lei da natureza e a lei civil contêm-se uma à outra e são de idêntica extensão... As leis da natureza não são propriamente leis... Só depois de instituído o Estado elas efetivamente se tornam leis, nunca antes, pois passam então a ser ordens do Estado, portanto também leis civis, pois é o poder soberano que obriga os homens a obedecer-lhes". "Tudo o que escrevi neste tratado sobre as virtudes morais, e sua necessidade para a obtenção e preservação da paz, embora seja evidentemente verdadeiro não passa por isso a ser lei. Se o é, é porque em todos os Estados do mundo faz parte das leis civis. Embora seja naturalmente razoável, é graças ao poder soberano que é lei. Caso contrário, seria um grande erro chamar lei não escrita à lei de natureza, sobre a qual tantos volumes foram publicados, com tão grande número de contradições, uns dos outros, e de si mesmos". Respectivamente, 26, 138 e 143. Por isso, Hobbes pode ser considerado defensor de um tipo de "positivismo jurídico", ainda que não totalmente coincidente com o tipo defendido por Kelsen, por exemplo. Isto se entendemos que o há de comum nas diferentes versões deste positivismo é a negação de uma das seguintes teses, que definem o "jusnaturalismo": (i) existem princípios de justiça universalmente válidos independentemente de serem reconhecidos efetivamente por instituições sociais ou indivíduos; (ii) um sistema normativo não necessita, para ser qualificado de *direito* ou *lei civil*, basear-se em atos contingentes de comando ou reconhecimento por parte de instituições ou indivíduos com poderes de coerção autorizados pela vontade do soberano. Esta caracterização do "jusnaturalismo" está baseada em Carlos S. Nino, *Ética y derechos humanos*, pp. 15-16. O "positivismo" de Hobbes parece envolver uma intrincada relação: um preceito contido numa "lei da natureza" só é lei se coincide com a lei civil, e uma norma só é de *direito* se é também justificada por alguma "lei da natureza" (é necessário, mas não suficiente, que seja comandada pelo soberano).

### 5.12. O "dilema do egoísmo".

Nem sempre os comentadores atentam para a teoria hobbesiana como um todo. Alguns simplesmente ignoram a construção precisa de Hobbes das "condições de motivação" moral. É claro que Hobbes não pode ser considerado um defensor do chamado "egoísmo de ato" (*act-egoism*), se este for definido como o tipo de egoísmo que explica o comportamento de uma pessoa que busca maximizar seus ganhos (quando medidos em *utilidade*, para ser mais preciso), sem se deixar limitar por uma regra qualquer que permita interagir de modo "beneficemente coordenativo" (ver a definição geral de "prática beneficemente coordenativa", p. 124). Comportamentos egoístas limitados por este tipo de regra exemplificam casos do chamado "egoísmo de regra" (*rule-egoism*). Leslie Mulholland, após propor a distinção entre estes dois tipos de egoísmo, sustenta que o egoísmo de ato é irracional, justificando esta afirmação da seguinte forma:

*"O problema com o egoísmo de ato é que não pode realizar o seu próprio fim da felicidade. Não há nenhum traço nele que indique como organizar desejos e apetites [isto é - tal como entendo este comentador - do princípio da busca da maximização ilimitada da satisfação dos apetites e desejos, não se deriva necessariamente nenhuma regra para organizar e limitar tal satisfação; ou, dito de outra forma, qualquer regra pode ser arbitrariamente adotada]. O egoísta racional está fadado à insatisfação. Se é racional buscar a felicidade, então é irracional ser egoísta."<sup>253</sup>*

Na verdade, precisando o sentido desta passagem, Mulholland pretende concluir que "o egoísmo de regra é, estritamente falando, impossível".<sup>254</sup> Pois não parece haver, segue este autor, qualquer motivação egoísta para aceitar a obrigação (expressa por uma regra) de limitar a busca da maximização da utilidade. Mas o que me interessa é que Mulholland mesmo admite que Hobbes oferece uma solução egoísta para o problema do egoísmo, a saber, o conflito interno insolúvel entre razões para agir e a interação violentamente conflituosa entre egoístas:

<sup>253</sup> *Egoism and morality*, p. 544.

<sup>254</sup> *Ibid.*, p. 549.

*"Uma solução poderia ser mostrar que é racional para o egoísta de ato estabelecer uma ordem de preferências que lhe permita organizar e limitar seus vários desejos e apetites... Tomem, por exemplo, uma regra que visa a manutenção de um prazer maximizado para um período longo de vida"*<sup>255</sup>

Todavia, Mulholland vê a tentativa de solução de Hobbes - assim como para "qualquer outra tentativa, interna ao egoísmo de ato, de resolver o problema da incapacidade do egoísmo de ato de alcançar a felicidade" - como "arbitrária e irracional",<sup>256</sup> pois a regra oferecida por Hobbes não pode ser justificada por razões egoístas (quer dizer internas ao egoísmo de ato):

*"Por que alguém deve adotar uma tal regra em vez de uma que envolva buscar a máxima intensidade de prazer numa vida curta, especialmente porque nunca podemos saber com certeza quais contingências danosas a vida nos reservará?"*<sup>257</sup>

Penso, contudo, que esta interrogação cética não tem alcance para minar a teoria hobbesiana. Como vimos, Hobbes não tem a pretensão de derivar o desejo de autopreservação inteiramente de algum imperativo categórico, racional independentemente de qualquer motivação do agente (ou, como diriam alguns, *objetivamente racional*). As regras morais da teoria ética hobbesiana são racionais para indivíduos cujo auto-interesse é sempre de longo-prazo. Elas só têm força normativa para quem, motivado pelo desejo de autopreservação, compreende que o melhor meio de satisfazê-lo é empreender uma busca deliberada e racional dos seus fins particulares; quer dizer, é precisamente a busca orientada pelas leis da natureza a que tem o melhor resultado: a maximização da continuidade do prazer individual.

Assim, para concluir este ponto, quero tomar de empréstimo um comentário de James Sterba e sugerir a sua aplicação ao caso de Hobbes:

<sup>255</sup> *Op. cit.*, pp. 544-545. Vale lembrar aqui que David Gauthier, não por acaso considerado um neo-hobbesiano, adota uma solução do mesmo tipo: a maximização que está em jogo para um agente racional não é a maximização das vantagens em cada situação interativa particular, mas a maximização das vantagens nas situações interativas que um indivíduo enfrenta na sua vida como um todo (isto é uma paráfrase da expressão "*overall maximization*").

<sup>256</sup> *Op. cit.*, p. 545.

<sup>257</sup> *Idem.*

*"Uma defesa mais promissora ... [é a de] que o conflito aparente entre moralidade e auto-interesse tem como fonte nossa falha em compreender o que é de nosso interesse de longo prazo. Alguns filósofos [e eu o exemplifico com Hobbes] afirmam que muitas violações da moralidade tem sua origem na ação impulsiva, para a qual o remédio é a busca deliberada e racional do auto-interesse de longo prazo"<sup>258</sup>*

Lembremos que o remédio, segundo Hobbes, é mais amargo. É preciso apelar para diferentes sanções da sociedade para forçar muitos a agir de modo cooperativo com os demais, mesmo na suposição de que alguns dos cidadãos são capazes de bem deliberar e decidir fazer o que é de seu interesse de longo prazo. Cito uma passagem na qual Hobbes é direto quanto ao ponto:

*"Não ignorais a força de um apetite irregular de riquezas, poder e prazeres sensuais, como ela domina a mais forte razão, e é a raiz da desobediência, do assassinato, da fraude, da hipocrisia e de toda sorte de maus hábitos; e que as leis humanas, ainda que possam castigar os frutos desses apetites, que são as más ações, não podem, contudo, arrancar suas raízes, que se encontram no coração. ... Como pode qualquer lei garantir segurança a um homem frente a outro, sendo a maior parte dos homens tão irracionais e parciais consigo mesmos como o são, e não sendo as leis senão letra morta, que por si mesma não é capaz de forçar alguém a agir de forma diferente da que lhe apraz, nem de castigá-lo ou lhe causar dano quando tenha cometido um malefício?"<sup>259</sup>*

---

<sup>258</sup> *Op. cit.*, p.46.

<sup>259</sup> *A dialogue between a philosopher and a student of the common laws of England*, pp. 8 e 10 na trad. espanhola. Ver também Rawls, *A theory of justice*, p. 240.

### 5.13. Como pode ser sempre racional agir de modo justo?

Determinar se as leis da natureza são *racionalmente estáveis* significa determinar se a justiça (leia-se: o conjunto de ações individuais *regularmente* limitadas pelas leis da natureza) é sempre individualmente racional (i. e., se agindo assim, em interações com outras que também o fazem, cada pessoa faz o melhor para si mesma). E isto é precisamente o que Hobbes pretende concluir ao responder ao "tolo" (*the foole*), quem sustenta que nem sempre é racional aderir àquelas leis, pois, na vida cotidiana - onde o esquema de cooperação definido pelas leis da natureza está supostamente em equilíbrio estável, essencialmente graças à existência de um "poder comum a recear" -, algumas vezes lhe é mais vantajoso ignorá-las (não aderir ao esquema de cooperação referido), deixando que os outros cumpram-nas unilateralmente.<sup>260</sup>

*"Os tolos dizem em seu foro íntimo que a justiça é coisa que não existe, e às vezes dizem-no também com a língua, alegando com toda a seriedade que, estando a conservação e a satisfação de cada homem entregue a seu próprio cuidado, não pode haver razão para que cada um deixe de fazer o que supõe conduzir a esse fim, e também, portanto, que fazer ou deixar de fazer, cumprir ou deixar de cumprir os pactos não é contra a razão, nos casos em que contribui para o benefício próprio. Com isso eles não pretendem negar que existem pactos, e que umas vezes eles são desrespeitados e outras são cumpridos, e que seu desrespeito pode ser chamado injustiça, e sua observância justiça. Mas perguntam se a injustiça ... não poderá às vezes concordar com aquela mesma razão que dita a cada um seu próprio bem, sobretudo quando ela produz um benefício capaz de colocar um homem numa situação que lhe permita desprezar, não apenas os ultrajes e censuras, mas também o poder dos outros homens ... Todavia, este especioso raciocínio é falso"*<sup>261</sup>

Hobbes não está disposto a aceitar a conclusão do argumento defendido pelos "tolos" simplesmente porque não aceita as suas premissas. Primeiramente tentemos esquematizar o argumento dos "tolos", para, a seguir, ver como Hobbes contra-argumenta:

(i) cada homem tem a seguinte psicologia básica: preocupa-se somente com a conservação e satisfação de sua pessoa (i. e., preocupa-se somente com seu *auto-interesse*);

<sup>260</sup> A conclusão do "tolo" é uma negação da condição de estabilidade racional expressa no item *b* na p.124.

<sup>261</sup> *Leviathan*, 15, 72-73. Os tradutores brasileiros cometem aqui outro erro (sublinhado a seguir). Lê-se nesta tradução: "Mas perguntam se a justiça ... não poderá às vezes concordar ...".

(ii) cada homem busca fazer aquilo que toma pelo melhor meio de promover seu auto-interesse (ou "*benefício próprio*");

(iii) em toda e qualquer *interação* com os demais (este é o tipo de ação que interessa para o argumento), não é contra a razão que cada um faça aquilo que "supõe conduzir" à sua própria conservação e satisfação (idêntico a *De Cive* I, VII, 83: "(...) nada é feito contra o uso da reta razão, quando a conservação própria é buscada por todos os tipos de meios") [*versão fraca do egoísmo racional*];

(iv) é racional fazer aquilo que é mais conducente ao próprio bem (ou, que conduz ao maior bem, ao maior *benefício próprio*) [pela definição de *racionalidade* como *maximização de algo*; logo, aplicada à teoria ética em questão, como *maximização do bem*];

(v) se, numa dada circunstância, desrespeitar um pacto previamente acordado (o que pode ser descrito como "cometer uma injustiça") é a ação racional - a que maximiza o benefício daquele que o desrespeita -, então agir justamente nem sempre é "uma coisa boa" (i. e., agir assim pode resultar em dano; alguém pode fazer um mal para si ao agir deste modo);

(vi) logo, agir justamente não é por natureza um bem (nem sempre resulta em "benefício próprio" de quem assim o faz), nem, por isso mesmo, racional (i. e., "a justiça é coisa que não existe", se definimos justiça como a prescrição universal do "cumprimento de pactos" (em todas as circunstâncias onde, graças à existência de um "poder civil", há garantia de cumprimento de promessas mútuas).<sup>262</sup>

A estrutura argumentativa que Hobbes usa é uma aplicação de seu método filosófico: se alguém argumenta (i. e., raciocina ou calcula com o uso de palavras) partindo de definições (i. e., primeiro estabelecendo o significado de "nomes gerais" ou conceitos)<sup>263</sup> e deduzindo as conseqüências das asserções nelas contidas, então, caso

<sup>262</sup> Hobbes considera que o uso significativo dos termos "justo" e "injusto" (seu contrário) pressupõe a verdade da proposição "existe alguma espécie de poder coercitivo, capaz de obrigar igualmente os homens ao cumprimento de seus pactos, mediante o terror de algum castigo que seja superior ao benefício que esperam tirar do rompimento do pacto". *Leviathan*, 15, 71. Esta consideração foi acrescentada ao esquema acima porque Hobbes irá pressupô-la antes de argumentar contra a conclusão defendida pelos "tolos". Ao fazer isso, Hobbes está aplicando uma das regras fundamentais do seu método filosófico: "... é necessário a qualquer pessoa que aspire a um conhecimento verdadeiro examinar as definições dos primeiros autores, ou para corrigi-las, quando tiverem sido estabelecidas de maneira negligente, ou para apresentar as suas próprias. Pois os erros de definições se multiplicam à medida que o cálculo [das conseqüências dos nomes ou "apelações"] avança e conduzem os homens a absurdos, que finalmente descobrem, mas que não podem evitar senão calculando-os novamente desde o começo, no qual [todavia] reside a base [mesma] de seus erros". *Leviathan* 4, 15; alterei um pouco a tradução das últimas três linhas, pois creio que a tradução brasileira (*Os pensadores*) faz surgir na passagem uma ambigüidade não presente no texto original. No parágrafo imediatamente anterior ao recém citado, Hobbes diz o porquê desta regra metodológica.

<sup>263</sup> Isso deveria ser uma mudança para melhor da argumentação filosófica (o que, para Hobbes, quer dizer científica) com relação às disputas escolásticas, em muitas das quais "os disputantes [nem

estabeleça uma conclusão falsa, tal conclusão (que Hobbes denomina "inferência geral") não expressa propriamente um "erro" (i. e. uma proposição contingentemente falsa, quer dizer, cuja negação é possível), mas um "absurdo", quer dizer, uma contradição, porque logicamente impossível de ser verdadeira (ou, fazendo uma consideração extemporânea, uma "proposição [que] é falsa para todas as possibilidades de verdade").<sup>264</sup>

O que Hobbes pretende fazer é mostrar que se as definições usadas pelo "tolo" forem corrigidas, a sua conclusão é falsa; e se não o forem - se o "tolo" não se deixar persuadir da necessidade de corrigir o uso "negligente" que faz de certos termos -, a sua conclusão é uma afirmação contraditória, que não tem sentido, não podendo descrever nenhuma situação possível da realidade. Uma contradição é uma asserção por meio da qual "nada mais concebemos senão o som" da conexão de um nome a outro; é absurda, insignificante e sem sentido - porque não podemos saber qual é o caso (o que é *real*, diria Hobbes) se ela é verdadeira, já que ela nem sequer tem a possibilidade lógica de ser verdadeira.<sup>265</sup>

Primeiro, examinemos a "objeção semântica" de Hobbes. Ele sustenta que o argumento do "tolo" baseia-se numa má compreensão do uso significativo do termo "pacto". O "tolo" assume que "existem pactos" e também que o predicado "justo", quando aplicado às ações, significa "o cumprimento de um pacto". Hobbes constrói a sua objeção do seguinte modo. Para se dizer que algo seja um pacto, devem estar satisfeitas as seguintes condições de verdade:

a) diferentes indivíduos prometem-se mutuamente - buscando cada um algum *bem para si mesmo* - fazer algo (prestar algum serviço, transferir algum objeto, etc.) no qual o outro está interessado;

b) um indivíduo, ao prometer algo, está manifestando a sua vontade de abrir mão da sua liberdade ("direito de natureza") de deixar de prestar algum serviço, de não se desfazer de algum objeto, etc.;

c) por meio de uma promessa, alguém anuncia a outrem a sua concordância voluntária em renunciar a parte do seu *direito original*, isto é, parte da sua *liberdade*, a saber, aquela que tinha de obstaculizar tudo e todos visando garantir a sua própria conservação;

---

sequer se punham] de acordo quanto à significação dos termos de sua própria arte". *Leviathan*, 14, 68.

<sup>264</sup> Wittgenstein, *Tractatus*, 4.46.

<sup>265</sup> A primeira parte desta última afirmação está de acordo com *Leviathan*, 5, 19 e 21; a segunda, usa quase literalmente o aforisma 4.024 do *Tractatus*.

d) "o motivo e fim devido ao qual se introduz esta renúncia e transferência do direito não é mais que a segurança da pessoa de cada um, quanto a sua vida e quanto aos meios de preservá-la de maneira tal que não acabe por dela se cansar".<sup>266</sup>

e) a obrigação de cumprir uma promessa (e o pacto que ela expressa) não depende da promessa em si ("a força das palavras ... é demasiado fraca para obrigar os homens a cumprirem seus pactos"),<sup>267</sup> mas deriva a sua força motivacional do fato de que quem promete:

e.1) sabe que é possível cumprir a sua parte;

e.2) sabe que o outro confia que ele cumprirá a sua parte (não tendo qualquer evidência razoável para suspeitar disso);

e.3) sabe que, cumprindo primeiro a sua parte, pode esperar que o outro também não falte à palavra dada;

f) a força motivacional só é suficiente "para levar os homens a cumprir suas promessas" se cada um deles temer as conseqüências nocivas da ruptura unilateral do pacto;<sup>268</sup>

g) cada contratante teme vir a sofrer uma perda maior que o ganho que teria por romper o pacto unilateralmente quando sabe existir acima de si e dos outros um poder comum capaz de coagi-los a cumprir o contratado sob pena de infligir-lhes um dano que, fazendo um balanço, tornaria o rompimento desvantajoso.

A "objeção semântica" de Hobbes pode agora ser entendida. Para que algo seja um pacto, as promessas aí envolvidas são feitas por contratantes que pesam, nos seus cálculos auto-interessados, as conseqüências da existência de um poder coercitivo comum. Assim, o argumento do "tolo" não usa significativamente o termo "pacto" simplesmente por não se poder aplicar o termo "promessa mútua" às palavras proferidas em contextos aos quais não se aplica a condição g, segundo a qual "não pode tratar-se de promessas mútuas quando de ambos os lados não há garantia de cumprimento, como quando não há um poder civil estabelecido acima dos autores das promessas". As promessas porventura envolvidas

<sup>266</sup> *Leviathan*, 14, 66. Esta condição já estava brevemente enunciada no item *a*. Neste ponto, podemos observar que o fim e a motivação de toda renúncia e transferência do direito somente ajuda a confirmar a adoção, por Hobbes, de versões do egoísmo racional e do egoísmo psicológico respectivamente.

<sup>267</sup> *Leviathan*, 14, 70.

<sup>268</sup> *Quer dizer, as futuras sanções que lhe serão impostas.* Neste contexto, vale a pena lembrar a definição de "sanção" apresentada por Bentham:

"*Sanctio*, em latim, era usada para significar o ato de ligar e, por efeito de uma transição gramatical comum, tudo o que serve para ligar uma pessoa - isto é, ligá-la à observância deste ou daquele modo de conduta (...)

Uma sanção é uma fonte de ... *dores* ou *prazeres*; os quais, conforme estiverem relacionados com estes ou aqueles modos de conduta, atuam como *motivos*, sendo na verdade as únicas coisas que podem atuar como tais." *Uma introdução aos princípios da moral e da legislação*, cap. III, nota 6.

faço a vista  
de vista



na situação descrita não satisfazem as condições que autorizariam tal situação ser também descrita como um pacto.<sup>269</sup>

Podemos agora supor que o "tolo" aceita a correção das definições. Neste caso, Hobbes pretende mostrar que a conclusão do seu raciocínio é falsa. Podemos dividir a conclusão em duas afirmações:

"Violar a fé" - ou seja, deixar de cumprir sua parte num pacto depois do outro contratante já ter cumprido a parte dele - é, algumas vezes, a ação:

- i) *virtuosa*, isto é, aquela pela qual melhor se realiza o bem individual.
- ii) *racional*, isto é, aquela pela qual se obtém mais vantagens individuais;

Esta dupla afirmação é a mesma que Sócrates, no final do livro I da *República*, reconhece ter sido sustentada por Trasímaco. Sócrates resume a posição sustentada por Trasímaco como expressando uma dupla objeção:

i) não é verdade que os justos (i. e., que voluntariamente praticam sempre ações justas porque consideram-nas um bem em si mesmas) tenham uma vida melhor e sejam mais felizes que os injustos (i. e., que voluntariamente praticam às vezes ações justas não

<sup>269</sup> Esta última afirmação é uma paráfrase explicativa da frase concisa de Hobbes: "... essas promessas não são pactos". *Leviathan*, 15, 73. Poderíamos repassar o problema nuclear do que chamei de "objeção semântica" por meio de uma apresentação em termos mais contemporâneos. Assim, diríamos: fazer um proferimento do tipo "prometo que ...", afirmar estas palavras, não é suficiente para realizar a ação de prometer. Pode simplesmente se tratar de um "proferimento enganador ou fraudulento".

Numa "falsa" promessa, falta um componente elementar: a intenção de fazer o que é descrito pela expressão que complementa "prometo que ...". Isto porque o ato de prometer não se constitui somente em proferir tais palavras. Deixa-se de prometer - é um *ato malogrado* - quando se tenta fazê-lo sem que certas condições apropriadas estejam satisfeitas. Dizendo-as insinceramente, o que se diz não chega a ser uma promessa. Austin elucidou o ponto:

"A insinceridade de uma asserção é a mesma que a de uma promessa. 'Prometo, mas não tenho a intenção de cumprir o prometido' é paralelo a 'isto é assim, mas eu não o creio'. Dizer 'prometo' sem ter a intenção correspondente é análogo a dizer 'isto é assim' sem se crer de fato no que se diz". *How to do things with words*; p. 55 na trad. brasileira.

Que *A* está realmente prometendo fazer *x* ao dizer "prometo fazer *x*" *pressupõe* a verdade da sentença declarativa "*A* tem a intenção de fazer *x*". Uma observação adicional resulta do emprego de uma analogia entre os diferentes propósitos dos usos de sentenças. Enquanto que o propósito do uso de uma *declaração (statement)* é indicar algum aspecto da realidade, o propósito da *sentença performativa* "prometo fazer *x*" é *comprometer-se a fazer x*, manifestar a outrem a intenção de, nas circunstâncias apropriadas, não deixar de fazer *x*. Recorro novamente a Austin:

"Assim como o propósito de uma asserção se frustra devido a uma contradição interna, ... o propósito de um contrato também se frustra se disser 'prometo, mas não devo fazer o prometido'. Esta expressão me compromete, mas ao mesmo tempo, anula o compromisso. Trata-se de um procedimento que anula a si próprio". *Ibid.*, p. 56. [ACRES.]

porque queiram ser justos, mas porque parecer justo lhes trará vantagens ou benefícios próprios);<sup>270</sup>

ii) não é verdade que a ação justa é, em todas as circunstâncias, mais vantajosa que a injusta.<sup>271</sup>

Como Hobbes bem observa, "um homem iníquo [i. e., injusto] [não] deixa de ser assim considerado por causa das ações que faz ou deixa de fazer devido ao medo, pois sua vontade não é determinada pela justiça, mas pelo benefício aparente do que faz".<sup>272</sup> Ou ainda: "Um homem justo é o que tem vontade constante de viver justamente. Se pedirdes mais, duvido que algum homem vivo possa estar compreendido na definição".<sup>273</sup> Mas Hobbes não pode endossar a afirmação de que os justos consideram as ações justas *um bem em si mesmas*, se com esta expressão indica-se uma *razão para agir* que exclui inclusive as chamadas "*consequências intrínsecas*" (ver nota 15) das ações justas.

Numa comunidade, nem todos indivíduos que fazem o que é justo fazer são *homens justos*. Justos são aqueles que seguem as prescrições das leis porque reconhecem que é o *melhor* que podem fazer por si mesmos. Não tomam cada ação particular de obediência como um mal necessário, uma alternativa menos atrativa a escolher sempre que for *a melhor possível*, dadas as limitações das circunstâncias (entre estas incluídos, principalmente, o risco de ser pego violando a lei e, a ele associado, o *medo* de vir a sofrer um mal maior do que aquele a que se sente constrangido a sofrer ao cumprir a lei). Homens justos não lamentam, como o fazem os que não têm *vontade constante de viver justamente*, qualquer perda por não serem livres para "fraudar" os demais sem impunidade; não o fazem porque desejam buscar prazer seguindo o que é correto fazer.<sup>274</sup> Poderíamos ainda

<sup>270</sup> Ver *República*, 352d, 359b e 361b-c.

<sup>271</sup> Ver *República* 354a-c. As objeções de Trasímaco são as mesmas objeções do "tolo". Se Hobbes de fato apresenta um bom contra-argumento, é então capaz de refutar ambas objeções simultaneamente.

Como aponta Hobbes, o "tolo" sustenta que agir injustamente é racional (maximiza o benefício próprio) naquelas circunstâncias nas quais "é impossível que daí resulte qualquer dano", visto que nelas o "homem iníquo" coloca-se ao abrigo não só dos "ultrajes e censuras", como também das possíveis represálias do poder dos outros homens. Cf. *Leviathan*, 15, 72.

<sup>272</sup> *Leviathan*, 15, 74.

<sup>273</sup> *Diálogo entre un filósofo y un jurista*, p. 27.

<sup>274</sup> Ver, por exemplo, *De cive*, III, V.

Aristóteles apontou o erro das concepções que procuram definir as virtudes como "estados de impassibilidade e repouso": se uma virtude é uma disposição para agir, e uma ação é medida com relação ao prazer que a acompanha (ou ao desprazer ou sofrimento que não a acompanha), uma virtude se define por um certo modo de promover o prazer (ou evitar o desprazer). Uma virtude

perguntar: é possível ser justo sem ser prudente? Segundo Hobbes, parece possível *ser prudente sem ser justo*; isto poderia ser explicado da seguinte forma: é possível que agir fazendo a ação justa seja, numa dada circunstância, uma mera coincidência para quem escolhe agir prudentemente.

Resta um ponto importante a esclarecer: o homem justo e o injusto parecem ser ambos *hedonistas*. Ainda assim, deveria haver diferença entre seus *hedonismos*. Seria apenas uma diferença quantitativa, quer dizer, agir de modo justo é apenas a estratégia mais eficaz para maximizar as chances de obter prazer a longo prazo (maximizar, neste período de tempo, a *utilidade esperada*)? Mas o interesse pelo “longo prazo” de sua vida não seria um fator a qualificar diferentemente o *hedonismo* do homem justo? Se houver então uma diferença qualitativa, poderia ela justamente repousar na existência de uma *virtude fundamental*, “fonte” das outras virtudes, a *justiça* entre elas? Não saberia dizer se, para Hobbes, existe uma tal virtude (a *prudência*, talvez?).<sup>275</sup> O que se pode dizer é que o hedonismo do homem justo se qualifica especificamente: o justo deleita-se em (satisfaz o desejo que tem de) “agir com justiça”, “fazer aquilo que é justo”. Além disso, há uma pista plausível para responder aquela pergunta: o desejo de fazer o que é certo só é compatível (i. e., só pode ser realizado juntamente com) com o desejo de não se deixar vencer pelo prazer e o desejo de prolongar o prazer de sua vida maximizando a sua continuidade.

Aceita esta perspectiva, há uma consequência importante. O hedonismo defendido por Hobbes é de um tipo específico. É preciso analisar a afirmação feita em 4.7: bom é o que produz prazer. Esta equivalência define uma teoria hedonista qualquer.<sup>276</sup> Há dois usos possíveis do termo bom, compatíveis com a teoria ética hobbesiana: *x* (ou fazer *x*) é bom porque dá prazer e *x* (ou fazer *x*) é bom porque é um meio para obter mais prazer no

---

expressa um modo correto de guiar a busca do prazer; ou ainda, um modo correto de ponderar o prazer como elemento da deliberação e escolha. Ver *Ética a Nicômaco*, II, 2, 1104 b10 - 1105 a1.

<sup>275</sup> “Fonte” é um termo vago. Possibilidades da relação entre a prudência como virtude intelectual (a razão como cálculo é apenas um componente desta virtude) e as demais virtudes foram assinaladas por Aristóteles ao contrastar a sua própria concepção com a de Sócrates:

“(…) para a parte moral da alma, existem dois tipos de virtudes, a virtude natural e a virtude propriamente dita, e destas duas, a virtude propriamente dita não se produz sem ser acompanhada de prudência. Eis por que alguns sustentam que todas as virtudes são formas de prudência, e Sócrates, no seu método de investigação, tinha razão num sentido, mas errava num outro: ao pensar que todas as virtudes são formas da prudência, ele estava enganado, mas dizendo que elas pressupõem a prudência, estava inteiramente certo.” *Ética a Nicômaco*, VI, 13, 1144 b15-21.

<sup>276</sup> Cf. William Frankena, *Ethics*; p. 99 na trad. brasileira.

futuro (ou, porque dá *poder*).<sup>277</sup> O homem injusto motiva-se a buscar de forma *desregrada* o prazer (esta busca caracteriza-se como uma *incontinência* face aos prazeres). Desde o ponto de vista das virtudes morais, que é o que interessa a Hobbes,<sup>278</sup> o homem injusto erra no modo como busca aumentar o seu poder, porque a continuidade do prazer pode ser *racionalmente* assegurada, assegurada *de modo otimizado*.

Afirmar que, para Hobbes, a continuidade do prazer é o Bem não basta (ver esquema da p. 98). Platão criticou as teorias éticas que identificam o Bem a algo *simples*: à prudência, ou à razão ou ao prazer. O Bem é algo *composto*, uma combinação destes três elementos; o prazer é apenas um dos bens, sustentou Platão.<sup>279</sup> A resposta de Hobbes parece ser, neste nível de abstração, similar. Parafraseando Aristóteles, pode-se dizer: os prazeres são desejáveis, mas não às custas de qualquer ação. Assim, o homem justo - ao contrário do iníquo, que obedece às leis porque tem medo de sofrer maiores danos, de ser privado de outros prazeres que cobiça - tem sua vontade determinada por uma qualidade que o predispõe a *desprezar* “ficar devendo a satisfação da sua vida à fraude ou ao desrespeito pelas promessas”.<sup>280</sup>

<sup>277</sup> Usando a terminologia apresentada por Frankena: *x* é *inerentemente* bom e *x* é *instrumentalmente* bom, respectivamente. *Op. cit.*, pp. 96-97.

<sup>278</sup> “(...) as leis da natureza, que consistem na equidade, na justiça, na gratidão e outras virtudes morais destas dependentes ... [são] qualidades que predispõem os homens para a paz e a obediência”. “Tudo o que escrevi neste tratado sobre as virtudes morais, e sua necessidade para a obtenção e preservação da paz ...”. *Leviathan*, 26, 138 e 143.

<sup>279</sup> Ver *Filebo*, 22a-22e e Aristóteles, *Ética a Nicômaco*, X, 2, 1172 b 25-35.

<sup>280</sup> *Leviathan*, 15, 74. A tradução brasileira contém erros nesta passagem. A disposição de caráter que predispõe o homem justo a *querer agir justamente* e, por isso, tomar o maior cuidado para *escolher* ações justas é o que Hobbes denomina de “nobreza ou valentia da coragem” (*Noblenesse or Gallantnesse of courage*) [na tradução ininteligível da coleção “Os Pensadores”: “nobreza ou coragem”]. Prossegue Hobbes: *justiça* é o nome da virtude que se atribui à conduta do homem justo.

#### 5.14. Por que agir conforme as "leis da natureza"?

A fim de testar a maneira, aqui apresentada, de analisar a validade das "leis da natureza" - isto é, determinar que condições devem estar satisfeitas para que tais leis possam ser sustentadas como convenções racional e motivacionalmente estáveis - e saber se ela pode ser considerada como uma interpretação elucidadora do tratamento que Hobbes dá ao tema, cito um trecho importante do *De cive*:

*"É algo evidente que as ações dos homens procedem da sua vontade e que esta vontade é governada pela esperança e pelo medo, de forma que os homens facilmente tendem a violar as leis, todas as vezes que desta violação possam esperar obter um bem maior ou um mal menor. Deste modo, toda a esperança que alguém tem de estar em segurança e de estabelecer apropriadamente a sua conservação está fundada na força e na habilidade, pelas quais ele espera prevenir-se dos desígnios do seu próximo e eludi-lo, o que prova que as leis da natureza não obrigam uma pessoa a observá-las assim que delas adquira conhecimento, como se elas lhe promettessem toda sorte de segurança. Porém, enquanto não tivermos garantia contra a invasão de outrem, devemos estar preparados para qualquer ataque e usufruir deste primeiro direito que a natureza nos dá sobre todas as coisas e que nos mantém no estado de guerra. Porque é suficiente a cada um, para levar a cabo a lei da natureza, que tenha uma disposição interior para a paz, quando não tiver qualquer suspeita de que aquela lei não será desobedecida pelos outros."<sup>281</sup>*

O problema genérico apontado por Hobbes pode ser descrito assim: nas interações de cada indivíduo com os demais, como ponderar as possibilidades de ação dos demais como fator decisivo na determinação dos seus próprios ganhos? Se estas interações se dão no "estado de natureza" ou em sociedade é algo que não muda o objetivo de um agente racional auto-interessado: ele prefere sempre um ganho maior a um menor de prazer. No "estado de natureza", pelo fato de não haver nenhuma garantia contra o ataque

<sup>281</sup> V, I, 136. A tradução desta passagem, feita pelo especialista em Hobbes Renato Janine Ribeiro, a partir da versão inglesa da obra, é um pouco diferente. Este tradutor afirma, referindo-se à tradução francesa, que "o resultado, na pena de Sorbière, é um texto muito elegante, de feitio clássico, e que arredonda as frases, explica-as, em alguns casos até as modifica - embora a única infidelidade maior a Hobbes esteja, tanto quanto pude perceber, no capítulo VIII, quando Sorbière traduz *servus* ou *servant* como *esclave*, o que implica, contra a expressa opinião do autor, dizer que o escravo firma um pacto com seu captor" (Apresentação à sua tradução do *De cive*, pp. XXXIII e XXXIV). Minha intenção aqui não é discutir esta tradução. Todavia, supondo-a fiel à versão inglesa, nota-se, nesta única passagem, algumas diferenças importantes; por isso mesmo, talvez seja mais adequado também fazer uso dela.

dos outros, cada agente adota a ação que considera a mais eficaz na produção de prazer (a que maximiza a sua expectativa de ganho - a utilidade esperada), independentemente do que os outros fizerem: "... a esperança que cada homem tem quanto a sua segurança e autoconservação consiste em que, pela força ou habilidade, ele possa lograr o seu próximo, seja abertamente, seja por algum estratagema".<sup>282</sup> O problema em ambos os estados (*de natureza e de sociedade*) é que cada agente não pode decidir racionalmente o que fazer sem se preocupar com o que os outros venham a fazer.

Este tipo de problema envolve o que os economistas denominam como *externalidades*. Quando uma pessoa age visando beneficiar-se, pode fazer algo que também afeta outras pessoas. Mas pode também não obter para si o que espera obter porque, ao agirem da mesma forma, outras pessoas fazem dela o alvo da soma de muitos *pequenos "efeitos colaterais"* (ou *externalidades*), que a afetam positiva ou negativamente. Ou seja, ao benefício primário, intencional, de cada agente, soma-se um benefício ou um dano cumulativos.

Caso o efeito secundário resultante seja benéfico, temos a operação de um mecanismo de "mão invisível", termo usado por Adam Smith para explicar como é possível a perseguição do auto-interesse servir ao interesse comum (enquanto pessoas racionais, isto é, que preferem sempre maiores a menores benefícios): cada pessoa faz melhor por si mesma (garante da melhor forma o seu bem-estar futuro) sempre que faz no momento o que lhe parecer melhor fazer (dito de outra forma, o que é racional para ela agora fazer identifica-se com o que é do seu interesse de longo prazo).<sup>283</sup>

No caso do somatório das externalidades ser um efeito danoso, piorando a situação para todos, o mecanismo pode receber o nome de "*contrafinalidade*" (o termo foi primeiramente usado por Sartre). Um exemplo simples de *contrafinalidade* é o seguinte: num jogo de futebol, cada espectador fica em pé visando ter uma visão melhor dos lances importantes; porém, o resultado é que ninguém consegue ver o que acontece, além de todos se cansarem de ficar em pé (pensemos na situação dos que estão na primeira fila; ainda que possam ver o que esteja acontecendo, também se cansarão, e provavelmente serão os mais

<sup>282</sup> *De cive*, V, I; tradução de Renato Janine Ribeiro.

<sup>283</sup> "Futuro" e "longo prazo" de fato referem-se aqui também ao futuro próximo ou prazo médio; sem incorrer em disputas verbais, o que se pretende é diferenciar conceitualmente estes casos daqueles expressando uma *atitude imediatista* frente ao tempo.

agredidos - pelo menos verbalmente - por serem identificados como os desencadeadores do efeito dominó).

Cada agente concebe a *eficácia* da lei da natureza (i. e., da eficácia do comportamento pelo qual tal lei é cumprida ou levada a cabo) para a produção da paz como dependente não só do próprio esforço (resultado da sua "disposição interior ou mental" de produzir um estado no qual está garantida a segurança da sua pessoa),<sup>284</sup> mas do esforço dos outros. A produção da paz é o resultado de um comportamento *instrumental*. Instrumental porque quem o adota está interessado no resultado alcançável por seu intermédio. Além disso, o agente hobbesiano, ao motivar-se pelos resultados efetivos possíveis de suas ações, leva em conta somente o que cada um deles "promete" para o seu próprio bem-estar. Estas duas condições nos autorizam a chamar este tipo de pessoa de *egoísta*.<sup>285</sup> Mas também é do interesse do agente hobbesiano (o tipo de pessoa que a teoria considera suscetível de influenciar-se por considerações que são, segundo a mesma teoria, éticas; ou, dito de outra forma, o tipo de pessoa em cuja vida as considerações éticas têm força)<sup>286</sup> que a maioria das outras pessoas faça o mesmo que ele faz: agir conforme as leis da natureza.

"Agir conforme as leis da natureza" é a descrição de ações individuais cujas explicações pressupõem: (1) querer fazer o que uma lei da natureza ordena; e (2) ter uma razão - i. e., desejar obter algo que se crê ser uma consequência da ação prescrita (ou desejar evitar algo que se crê ser uma consequência da omissão da ação prescrita) - para querer (ou estar motivado a) fazer o que tal lei ordena.<sup>287</sup> Dada esta estrutura explicativa, são possíveis diferentes motivações para fazer o que as leis da natureza ordenam.

Considerando-se a forma geral das "leis da natureza" ou, tal como Hobbes a denomina, "a lei primeira e fundamental da natureza", há um modo trivial de satisfazer as duas pressuposições acima: toda pessoa que "age conforme as leis da natureza" faz o que uma lei da natureza qualquer ordena porque reconhece que a ação aí prescrita é um meio

<sup>284</sup> Ver, por exemplo, *Leviathan*, 14, 66, onde Hobbes aponta em que consiste tal "segurança".

<sup>285</sup> Nem toda pessoa preocupada com resultados é egoísta; um utilitarista, sabendo que muitas pessoas de algum modo fraudam o imposto de renda (declarando suas rendas parcialmente e deduzindo-as incorretamente), não sonega o seu imposto, porque este, justamente nestas circunstâncias, se faz mais necessário para se obter uma maior "felicidade do todo" (entenda-se aqui, p. ex., maiores benefícios em termos de serviços públicos disponíveis para toda a população).

<sup>286</sup> Ver 1.1.

<sup>287</sup> Ver 1.3.

que contribui para a realização do estado de coisas que deseja, a paz (e evitar algo que não deseja, a morte violenta), na verdade, um estado de coisas constituído por milhares de tarefas só realizáveis cooperativamente, isto é, tarefas interligadas ou interdependentes - <sup>alt.</sup> [quer dizer, nas quais as pessoas em interação baseiam, para conseguir benefícios capazes de garantir a sua própria sobrevivência, na existência da *reciprocidade*.]

A reciprocidade está indicada no *imperativo condicional* que expressa a forma geral das "leis da natureza" (ver 5.6). Assumindo que a obediência ao imperativo "faça x" ocorre se a *sentença de obediência* "x é feito", que descreve a *ação correspondente*, é verdadeira; para o imperativo das "leis da natureza", se a sentença de obediência é falseada, i. e., se o agente deixa de fazer x, há duas alternativas: o imperativo não é desobedecido se o agente a quem está endereçado deixa de fazer x porque, nas circunstâncias em que se encontra, não considera razoável esperar reciprocidade; mas é desobedecido se tal agente deixa de fazer x considerando razoável esperar reciprocidade.<sup>288</sup>

Há fatores decisivos para formar o juízo sobre a "razoabilidade" da cooperação em dadas circunstâncias. Supõe-se que é do interesse de longo prazo de cada indivíduo cooperar com os demais justamente porque, na existência de sanções sociais formais - tipos variados de penalizações - (e mesmo informais, tais como o ostracismo social), ele pode esperar que seu comportamento seja reciprocado. É claro que esta esperança baseia-se na hipótese de que os demais guiam-se (são motivados) fundamentalmente (na maioria das circunstâncias) pelo seu auto-interesse de longo prazo.

Assim, se alguém for incapaz de motivar-se pelos ganhos futuros, ou se tal motivação for superada pelo "perverso desejo de vantagens imediatas", [ainda que possa conhecer a "norma de reciprocidade", não se sentirá obrigado a cumpri-la. Como qualquer outra norma, se não pode ingressar como "causa" de uma ação, também não pode ser aceita <sup>alt.</sup> como prescrição; neste caso, como o diz Hobbes, não se trata propriamente de uma lei, quer dizer, um ditado que se mostra como obrigatório, que se tem obrigação de observar.]<sup>289</sup>

<sup>288</sup> Bernard Williams discute de modo refinado as possíveis análises da forma lógica de um *imperativo condicional*; principalmente - o que é bastante relevante para a presente discussão - a não-equivalência entre um *imperativo condicional* e o imperativo de um *condicional material*. Ver, por exemplo, *Problems of the self*, pp. 187-191. A melhor apresentação que conheço da distinção entre *condicional material* e *condicional formal* (o imperativo que expressa a forma geral das "leis da natureza" pertence a uma de suas subclasses) é a feita por Tarski, no parágrafo 8 de *Introdução à lógica*.

<sup>289</sup> Ver *De cive*, I, III, XXVI-XXVII.



### 5.15. A disposição de "abraçar a paz quando ela se mostrar viável"

Assim, a pergunta que cabe é a seguinte: em que circunstâncias a paz pode "se mostrar viável", ou, para dizê-lo de modo mais articulado, em que circunstâncias cada um pode razoavelmente esperar que seu esforço para produzir a paz seja compartilhado pelos demais? Se considerar os demais como agentes inteligentes, também os conceberá como preocupados com o que ele próprio venha a fazer. Para todo agente vale, então, o mesmo raciocínio: a minha disposição para cumprir a lei da natureza (por querer - estar motivado a - produzir a paz) só é um motivo suficiente para cumpri-la se eu estiver certo que os demais também decidem cumpri-la.

Ora, a decisão de cumpri-la ou não é, respectivamente, o último apetite ou aversão da deliberação ou raciocínio. Em todo caso, o apetite ou aversão dependem da crença que cada um tem de conseguir garantir a sua própria preservação. Quanto maior o grau da crença<sup>290</sup> que se tem em obter a autoconservação, maior o apetite por fazer a ação que a tem como consequência. E maior a esperança de obtê-la ("esperança" é, para Hobbes, o nome que se dá a um apetite, sempre que o consideramos como relativo à "opinião que os homens têm da probabilidade [*likelihood*] de conseguirem o que desejam").<sup>291</sup>

<sup>290</sup> Se analisamos o conceito de crença de tal modo que só faça sentido usá-lo para afirmar de um sujeito *S* que "*S* crê que tal-e-tal é o caso" (ou a forma equivalente: "*S* não crê que tal-e-tal não é o caso") ou "*S* não crê que tal-e-tal é o caso" (ou a forma equivalente: "*S* crê que tal-e-tal não é o caso") e não faça sentido dizer que "*S* crê *mais* que tal-e-tal é o caso do que crê que tal-e-tal não é o caso", então a noção de "graus de crença" é absurda. Pode-se, no entanto, admitir o uso plausível desta noção em algumas circunstâncias. Por exemplo, quando quisermos dizer que *S* tem diferentes expectativas quanto às ocorrências futuras de dois eventos diferentes; podemos dizer, por exemplo, "*S* crê mais intensamente que choverá amanhã do que (crê que) o mundo terminará amanhã". Uma forma equivalente e sem tanta conotação psicológica de afirmar o mesmo é: "choverá amanhã" expressa um grau de crença maior que o expressado por "o mundo terminará amanhã".

A "probabilidade subjetiva" é justamente a medida numérica das crenças. Esta é, em síntese, a concepção da chamada "teoria clássica" sobre a natureza da probabilidade: esta expressa um grau (ou medida) da crença racional. Explicando melhor: "a probabilidade de um acontecimento que se vaticine depende do grau em que se creia racionalmente que ele pode ocorrer". Irving Copi, *Introdução à lógica*, p. 424. Trata-se de uma "crença racional parcial", uma crença baseada em informações limitadas (o que supomos conhecer e que chamamos de *evidências*) e cujo "peso" maior ou menor (para usar uma expressão de Keynes) é sempre *relativo* (nenhum acontecimento tem uma probabilidade *intrínseca*), isto é, "as crenças de um homem razoável mudam com o estado de seus conhecimentos". *Ibid.*, p. 426.

<sup>291</sup> *Leviathan*, 6, 25.

A "disposição para a paz" é um requisito essencial para garantir a estabilidade motivacional das leis da natureza. É precisamente a falta desta disposição que caracteriza a essência dos "tempos de guerra"; a guerra não se constitui somente por batalhas intermitentes, mas inclui também os lapsos de tempo "durante [os quais] a vontade de travar batalha é suficientemente conhecida".<sup>292</sup> Por isso Hobbes afirma que "a natureza da guerra não consiste na luta real, mas na conhecida disposição para tal".<sup>293</sup> Todavia, quando os homens têm a principal garantia que podem ter da realização da paz - o fato de terem instituído "um poder comum" capaz de "manter a todos em respeito" - e, ainda assim, têm sua vontade guiada por uma "aversão pela paz", são desfeitos aqueles vínculos interativos cuja continuidade determina o "cimento da sociedade". Este traço nuclear da natureza humana, que podemos descrever como a disposição de "abraçar a paz quando ela se mostrar viável",<sup>294</sup> é um pressuposto necessário da adesão individual às leis da natureza.

Em sociedade, esta disposição é genericamente caracterizada como a *disposição para agir justamente*, quer dizer, para cumprir com sua parte nos pactos celebrados. Na falta generalizada desta disposição, nenhuma sociedade pode subsistir. Este é o núcleo de uma observação "perfeita e clássica"<sup>295</sup> de Adam Smith em *The theory of moral sentiments*:

*"A sociedade, entretanto, não pode subsistir entre aqueles que estão prontos para se ferir e injuriar uns aos outros. No momento em que principia a injúria, no momento em que surgem o ressentimento e a animosidade mútuos, todos os vínculos se estilhaçam, e os diferentes membros de que ela consiste são, por assim dizer, dissipados e dispersados pela violência e oposição de seus afetos discordantes. Se há alguma sociedade entre ladrões e criminosos, eles devem, conforme o que é observado pelo senso comum, pelo menos abster-se de roubar e assassinar uns aos outros. A beneficência, portanto, é menos essencial para a existência da sociedade do que a justiça. A sociedade pode subsistir, ainda que não sendo no estado de maior conforto, sem a beneficência; mas, a predominância da injustiça necessariamente a destrói por completo"*<sup>296</sup>

<sup>292</sup> *Leviathan*, 13, 62.

<sup>293</sup> *Idem*.

<sup>294</sup> *De cive*, V, I; trad. de Renato Janine Ribeiro.

<sup>295</sup> Ver comentário de Keneth Lux em *O erro de Adam Smith: de como um filósofo moral inventou a economia e pôs fim à moralidade*, cap. 6.

<sup>296</sup> Parte 2, seção 2, cap. 3; p. 86.

### 5.16. Sobre a obrigação de observar as leis da natureza.

Num trecho citado mais acima (p. 143), Hobbes chega à conclusão de que *o que é moralmente correto fazer* não é exigido unilateralmente de cada agente que interage com outros, já que isto nem sempre é racional para agentes primordialmente interessados na sua própria conservação. Enquanto prevalecer o *estado de natureza*, a obrigação de observar as leis da natureza não é justificável racionalmente, uma vez que não estão dadas as condições que tornam tais leis eficazes.

Quanto a este ponto, Thomas Nagel parece ter se equivocado, ao ser influenciado pela concepção de moralidade como um sistema de regras estabelecendo a obrigatoriedade de determinadas ações (*fundamentalmente* motivadas por algum tipo de consideração pelos interesses, desejos, sentimentos ou bem-estar *dos outros*) mesmo quando realizá-las (i. e., obedecer tais regras) seja contrário ao auto-interesse de quem as realiza. Nagel conclui o seu já clássico artigo afirmando, entre outras coisas, que:

*"[As leis da natureza] contêm qualificações para garantir que em nenhum caso a adesão a elas seja contrária ao auto-interesse do indivíduo particular. Isto é, não somente é o caso que se todos os homens seguirem estas regras, o resultado será para o bem geral, como também é o caso que um único homem ao segui-las, mesmo se ninguém mais o fizer, não estará colocando em risco o seu próprio bem-estar"*<sup>297</sup>

Segundo Hobbes, não há outro modo de estabelecer a eficácia das leis da natureza a não ser a "saída" dos homens da sua "condição natural" para uma sociedade política. Para dizê-lo de modo mais preciso, o problema da determinação das condições

<sup>297</sup> *Op. cit.*, pp. 82-83; grifo meu. Mais tarde, Nagel reconheceu que a adesão às leis da natureza só é racional, para todos indivíduos auto-interessados, se também as circunstâncias desta adesão são tais que garantem que a preservação de *cada um* deles seja de fato realizada. Isto quer dizer, não se pode excluir das condições de validade dos princípios normativos da teoria ética de Hobbes - que devem estar satisfeitas para que a obrigação de obedecê-los possa ser racionalmente justificada - as chamadas condições de eficácia. Baseio este breve comentário na seguinte passagem:

"Subjective reasons hold for everyone, but assign in the first instance a different but related goal to each person: *his* happiness, *his* security, ... However, for a given subjective reason there may be certain circumstances which realize the relevant goal for everyone. Hobbes's ethical doctrine depends on the assumption that a stable, well-policed society will be overwhelmingly in the interest of each of its members, so that a mere consideration of his own individual interest provides each with a reason to pursue and preserve that condition". *The possibility of altruism*, pp. 95-96.

necessárias para garantir a *estabilidade motivacional* é o problema nuclear da teoria política de Hobbes. Isto está indicado numa conhecida passagem do *De cive*:

*“A maior parte daqueles que escreveram alguma coisa a propósito das repúblicas ou supõe, ou nos pede ou requer que acreditemos que o homem é uma criatura que nasce apta para a sociedade. Os gregos chamam-no zoon politikon; e sobre este alicerce eles erigem a doutrina da sociedade civil como se, para se preservar a paz e o governo da humanidade, nada mais fosse necessário do que os homens concordarem em firmar certas convenções e condições em comum, que eles próprios chamariam, então, leis. Axioma este que, embora acolhido pela maior parte, é contudo sem dúvida falso - um erro que procede de considerarmos a natureza humana muito superficialmente.”<sup>298</sup>*

A teoria ética de Hobbes deve, contudo, antes (i. e., na ordem lógica) mostrar a possibilidade da passagem de um estado a outro, e para isso é central a demonstração de que é racional para os contratantes abrir mão - isto é, aceitar uma restrição ou limitação - de parte da sua "liberdade natural" de usar todos os meios que consideram necessários para obter o máximo de satisfação dos seus próprios desejos. A teoria ética deve poder demonstrar a estabilidade racional das convenções limitando a "liberdade natural" de cada indivíduo.

<sup>298</sup> I, II; tradução de Renato Janine Ribeiro. Acontece que, como acrescenta Hobbes numa nota à passagem citada, as sociedades civis não são meras *reuniões* de homens naturalmente motivados a obedecer (concepção erigida sobre o pressuposto de que “o homem é uma criatura que nasce apta para a sociedade”) regras para a convivência pacífica, mas sim “obrigações” (*bonds*) constituídas por pactos entre indivíduos cuja preferência pelo cumprimento das regras acordadas depende da existência de condições externas - não somente da sua própria vontade - capazes de induzir a adesão dos demais indivíduos às mesmas regras. Ver a definição de “convenção” apresentada na nota 231. No prefácio do *De cive*, para citar um exemplo, esta concepção aparece com clareza: “(...) as disposições dos homens naturalmente são tais que, a menos que sejam restringidos pelo temor a algum poder coercitivo, todo homem sentirá desconfiança e temor de qualquer outro.”

### 5.17. Como é possível a cooperação entre egoístas?<sup>299</sup>

A concepção hobbesiana de moralidade pressupõe a introdução de um ingrediente essencial: uma restrição ou limitação racional sobre o que, *sob outras condições*, também é racional fazer, a saber, o "direito de natureza" de cada indivíduo fazer ilimitadamente tudo o que bem entender visando a maximização da vantagem individual. Poderia assim seguir a seguinte máxima: "tirar o máximo de vantagem *em tudo*", i. e., em todas as suas interações diretas com os outros e mesmo na ausência destes, pois se "ninguém o vê", poderia raciocinar esse indivíduo, não há razão para acatar qualquer norma que implique em desvantagens para si mesmo.

Ora, se concebemos o sentido de *racional* como a maximização da vantagem ou bem-estar individual, gera-se o seguinte paradoxo. Se de fato obedecer às "leis da natureza" implica em fazer restrições ao comportamento gerado pela busca da maximização da vantagem pessoal, então o comportamento moral não é mais o comportamento que a razão prescreve como o mais vantajoso. Por outro lado, se o comportamento mais vantajoso é aquele prescrito pelas "leis da natureza", então por que falamos de uma limitação ou restrição do padrão comportamental anterior? Poderíamos conceber o papel das convenções morais como simplesmente o de garantir, na prática, uma *coordenação dos interesses* dos agentes que interagem de um modo harmônico, quer dizer, de um modo tal que não seja necessário a ninguém se impor limitações na busca da maximização da satisfação pessoal, tal como ocorre na idealização usada na teoria econômica do *mercado perfeitamente competitivo*, isto é, um domínio de interações onde "a vantagem mútua é garantida pela atividade irrestrita de cada indivíduo na busca da sua própria maior satisfação, de tal modo que não há lugar, racionalmente, para restrições"<sup>300</sup>

<sup>299</sup> Esta seção é suplementada pelo apêndice II.

<sup>300</sup> Gauthier, *Morals by agreement*, p. 13. Creio que se pode responder afirmativamente a esta pergunta; porém, tal resposta precisa ser bastante cuidadosa para não obscurecer a relação conceitual adequada entre o contrato moral pressuposto no "contrato social" e o tipo de conduta que caracteriza agentes interagindo num "mercado perfeitamente competitivo". Esta relação foi assinalada com bastante clareza por Allan Bloom:

"(...) o mercado (a competição pacífica para a aquisição de bens) exige a prévia existência do contrato social (a concordância em submeter-se a contratos e a criação de um juiz para arbitrá-los e fazê-los cumprir), sem o que se instala o estado de guerra. Estar em guerra fazia parte da condição humana antes da existência da sociedade civil e é sempre possível retornar a ela. A força e a fraude [- na guerra, diz Hobbes, "as duas virtudes cardeais" -] ... não têm nada a ver com o mercado ["perfeitamente competitivo", penso que deveríamos precisar] e são ilegítimas dentro dele. A

Resumindo o paradoxo: estamos diante de uma conclusão expressa por uma disjunção inaceitável; a saber, ou bem não podemos justificar racionalmente a adesão individual às "leis da natureza", ou bem, se for justificável, tal adesão não pode implicar em qualquer limitação do chamado "comportamento racional individual".

Felizmente tal paradoxo pode ser desfeito se for possível provar que estamos diante de uma limitação comportamental<sup>301</sup> que é, ela mesma, *racional*. Mas como é possível sustentar esta afirmação?<sup>302</sup> No "estado de natureza", cada indivíduo, após ponderar sobre a completa insegurança no que diz respeito à continuidade de sua própria vida, procura, a cada momento, apropriar-se da maior quantidade de coisas que lhe possam garantir, por mais tempo, a satisfação dos seus desejos. Nesta última afirmação, estamos pressupondo que o desejo de autopreservação é fundamentalmente um desejo de continuar vivo durante o maior período possível. Além disso, que a *racionalidade destes* indivíduos se define como a capacidade de calcular o que deve ser feito para garantir a maximização da satisfação individual durante o maior período possível. Por isso, falamos de uma *racionalidade prudencial*; *prudência* tomada no mesmo sentido que Hobbes a toma: a habilidade de determinar, dentre as conseqüências das ações, também aquelas menos próximas.<sup>303</sup>

---

conduta racional dos homens em tempo de paz, no que a economia se especializa, não é a mesma que a sua conduta racional em tempo de guerra." *The closing of the american mind*, p. 374 na trad. brasileira.

<sup>301</sup> Esta limitação pode ser entendida como a limitação da liberdade que cada um tem, no estado de natureza, de adotar qualquer comportamento que creia como meio indispensável para a própria autoconservação.

A autoridade do soberano não é legítima para estabelecer a obrigação de um egoísta racional a fazer o que *não* quer fazer. Conceber as leis apenas como *limitações coercitivas* do comportamento é um traço da psicologia daqueles indivíduos que não são capazes de concebê-las como necessárias para a "estruturação da vida", pois, além de ajudá-los a não se prejudicar, permitindo, a cada um, que busque satisfazer os desejos que crê contribuir para a sua autopreservação, as leis criam condições sob as quais a ajuda mútua é racional. Desta ajuda depende a *autorealização* (cidadãos não são autosuficientes quanto aos meios que necessitam para a *satisfação de suas vontades e propósitos*). (Esta última afirmação está cf. *Leviathan*, 4, 13)

<sup>302</sup> Uma justificação mais precisa é oferecida por meio dos resultados teóricos da "solução do dilema do prisioneiro", pois tal como afirma Amartya Sen, o "jogo" do "dilema do prisioneiro" pode ser "tratado ... como o caso clássico do fracasso da racionalidade individualista". *Rational fools*, p. 340. Uma análise breve desta afirmação é oferecida pelo mesmo autor no artigo *Behaviour and the concept of preference*, in Elster (ed.) *Rational choice*, pp. 69-71.

<sup>303</sup> Uma análise adequada feita desde outro ângulo pode mostrar por que é possível dissolver o paradoxo em questão. Segundo Nicholas Rescher, é importante perceber que "o exemplo [da situação do dilema do prisioneiro] não mostra (como é freqüentemente dito) que a racionalidade pode ser prudencialmente contraprodutiva". *Rationality and moral obligation*, p. 40.

Também vimos que um raciocínio prudencial correto mostra a cada um que, no "estado de natureza", dependendo dos tipos de variáveis incluídos no cálculo realizado, ele tem duas estratégias alternativas possíveis para agir racionalmente:

1. é mais vantajoso *sempre* - portanto é uma estratégia dominante - agir, independentemente do que os outros fizerem, por meio da chamada "violência antecipatória";

2. se no raciocínio for incluído o estado que resultará da adoção, por parte de todos aqueles que interagem "comigo", da estratégia dominante descrita em 1, então para um número  $n$  de interações, é mais vantajoso cooperar, sempre que estiver seguro que os outros farão o mesmo; e isto é válido ainda que "eu" possa reconhecer que, com um certo risco, agir não-cooperativamente, sob condições "favoráveis", possa vir a resultar num maior benefício pessoal, considerando-se as conseqüências imediatas de um número reduzido de interações.<sup>304</sup>

Podemos agora perguntar: qual destas duas estratégias é racional adotar? Esta questão não tem, contudo, uma única resposta, uma vez que é possível respondê-la de uma forma condicional: se o agente for movido de modo dominante pelo desejo  $x$ , então é racional para ele adotar a estratégia  $y$ . Supondo então, como Hobbes o faz, que o desejo de autopreservação é basicamente um desejo de poder continuar satisfazendo os desejos que continuarão a formar-se *no futuro*, ainda que não sejam *agora* desejados, então a estratégia descrita em 2 é a racional, isto é, aquela que melhor garantirá a satisfação daquele desejo.

Neste caso, um agente racional - e que não superestima os seus ganhos imediatos - introduz limitações no comportamento que visa a maximização da sua satisfação por meio da maximização das vantagens tiradas *no presente* (e futuro imediato) porque tal estratégia é a que lhe resulta ser mais vantajosa num período de vida que se estende a médio e longo prazo (i. e., *no futuro*). Mas isto não basta, porque, independentemente da sua psicologia, só pode razoavelmente considerar que lhe é mais vantajoso *cooperar* se puder antecipar - no seu raciocínio - com alguma segurança o fato de que os outros farão o mesmo.

Segundo Hobbes, existe somente uma maneira de evitar o *caos* que caracteriza o estado de natureza, isto é, um estado onde os planos de cada indivíduo para satisfazer

<sup>304</sup> Aceitando-se o "argumento regressivo" no chamado "dilema do prisioneiro iterado", aceita-se que  $n$  deve ser um número finito mas não determinado de vezes. Além disso, é preciso pressupor que  $n$  abrange um período de tempo suficientemente grande para tornar a não-cooperação uma ação cujos ganhos imediatos sejam menores que as perdas não imediatas.

seus desejos são, durante o tempo todo, frustrados.<sup>305</sup> A única maneira de evitar este caos é criar um mecanismo de imposição de comportamentos cooperativos. Dada a natureza humana e as condições de incerteza que compõem a "condição natural da humanidade", só o uso da força é capaz de induzir as pessoas a agir de certa maneira. "Força" - tomando emprestado uma excelente definição - significa "qualquer ação designada a tornar uma prática indesejável mais cara para aqueles que poderiam sentir-se tentados a empenhar-se nela".<sup>306</sup>

Interessa-me sublinhar o seguinte ponto: o "contrato moral" propriamente dito pode ser descrito como a livre concordância sobre a necessidade da criação de um juiz cujo poder torne satisfeitas as condições sob as quais é racional, para cada agente, submeter-se a todos os contratos previamente acordados. Genericamente, esta descrição é válida independentemente de qual o tipo de agente que se pressuponha. Para o caso específico de *egoístas racionais*, o poder capaz de satisfazer as condições sob as quais é racional, para cada agente, cooperar é o poder de infligir sofrimento e dor, impondo coercitivamente, por meio do medo do sofrimento, a contenção daquelas condutas que podem contribuir para desestabilizar as condições que melhor conduzem aos fins desejados por cada um.<sup>307</sup>

Desde a perspectiva da *teoria dos jogos*, deve haver um raciocínio que justifica, para cada parte contratante, chegar a um acordo coletivo fixando regras para resolver as disputas que possam surgir entre elas. Tal raciocínio envolve algumas características próprias. A deliberação individual é parte de um processo denominado *escolha estratégica*.

<sup>305</sup> A expressão "planos frustrados" refere-se também àqueles planos cujas ações, por conduzirem a objetivos que colidem com objetivos de outrem, são tomadas como ameaça mesmo antes de serem realizadas. Um plano de um aliado oportunista para induzir o desarmamento unilateral é tomado como uma ameaça pelo país cujo objetivo seja garantir a sua segurança de preferência sem onerar demasiadamente sua receita com gastos militares.

<sup>306</sup> Elster, *Nuts and bolts*; p. 175 na ed. brasileira. Hobbes refere-se ao uso da *força* quando trata da necessidade de "um poder coercitivo capaz de obrigar [isto é, forçar] igualmente os homens ao cumprimento de seus pactos, mediante o terror de algum castigo que seja superior ao benefício que esperam tirar do rompimento do pacto." *Leviathan*, 15, 71-72. Para a equivalência entre *obrigar* e *forçar a*, comparar esta passagem com *Leviathan*, 17, 85. Um acordo coletivo entre *egoístas racionais* só é uma razão suficiente para cada um agir do modo nele estipulado se houver garantia de que a adesão individual é sempre a *melhor ação*, a que implica em maior ganho a longo prazo.

<sup>307</sup> Neste caso, vale a consideração de que as condutas incompatíveis com o fim do egoísmo racional serão, em sociedade, "acomodadas", por meio de leis, à vontade dos contratantes *qua* *egoístas racionais*. A concepção de *lei* pressuposta aqui é apresentada por John Austin no clássico *The province of jurisprudence determined* (1832), conferência I. Ver comentário de David Lyons, *Ethics and the rule of law*; pp. 47-54 na trad. brasileira.



Nesta, o agente racional escolhe baseado num raciocínio no qual nem todas as características do seu "ambiente" são *parâmetros*, quer dizer, nem todas equivalem a valores constantes. O agente estrategicamente racional leva em consideração o fato de que o ambiente é "povoado" por outros atores, que ele próprio é parte do ambiente de escolha deles, que sabem do seu comportamento e que levam-no em conta nas suas expectativas.<sup>308</sup>

Como bem assinala Jon Elster, a consideração do comportamento estratégico permite sublinhar o "fato crucial de que os atores humanos não somente fazem as suas escolhas com base nas suas expectativas sobre o futuro, ... mas também *com base nas suas expectativas sobre as expectativas dos outros*".<sup>309</sup> Desta forma, um agente estrategicamente racional não espera poder alcançar um ganho máximo, procedendo unicamente ao ajuste optimal de meios a seus objetivos, independentemente do que os outros esperam e venham a fazer. O que este agente deve fazer para fazer o melhor por si mesmo é uma variável dependente das ações feitas pelos outros.

Assim, o agente estrategicamente racional inclui no seu cálculo as possíveis expectativas dos outros agentes sobre o seu próprio comportamento. Em princípio, este agente poderia buscar influenciar a escolha dos outros de diversas maneiras; por exemplo, pode procurar dar aos outros apenas uma impressão de credibilidade, a fim de conseguir deles o que deseja sem nenhuma contrapartida sua. Trata-se de uma vantagem limitada, porque depende que este agente não seja quem faz o primeiro movimento no jogo interativo. Se isso não for assim, ou se introduzimos o fator tempo no jogo (o que permitiria aos adversários conhecer a verdadeira estratégia daquele jogador), a trapaça pode deixar de funcionar com eficácia.

Uma outra estratégia poderia ser a de tentar limitar as alternativas de escolha do adversário de modo a produzir para si mesmo o resultado mais conveniente. É uma tentativa de restrição unilateral do comportamento (pelo uso da força, da ameaça, ou de vantagens iniciais de outra natureza). Todavia, se bem ponderadas, qualquer uma dessas estratégias mostra-se coletivamente irracional, se pessoas com objetivos diversos fazem escolhas, visando realizá-los, guiadas diretamente pela busca racional de seus próprios objetivos, tudo pode resultar numa situação na qual cada pessoa tem seus objetivos frustrados em maior grau que o produzido caso adotassem outra regra de escolha.

<sup>308</sup> Por isso é capaz de ver a si mesmo como um jogador entre outros.

<sup>309</sup> *Ulysses and the sirens*, p. 19.

Porém, um acordo no “estado de natureza” é uma exigência plausível porque, enquanto não for instituído o poder soberano, as ações guiadas pela busca direta da maximização da consecução de fins individuais não deixam de corresponder à melhor escolha que um indivíduo pode fazer numa circunstância qualquer dentro daquele estado. A estabilidade requerida pelo contrato coletivo entre egoístas é associada a um estado de coisas descrito como *equilíbrio a longo prazo*. É a *solução* para a situação de escolha coletiva descrita como “estado de natureza”, correspondendo ao melhor resultado que cada jogador pode obter para si (dados os seus interesses, principalmente o maior deles, a saber, que deseja o maior ganho possível *a longo prazo*) diante das limitações certamente oferecidas pelos interesses dos outros jogadores (conforme uma maior ou menor oposição).

Mas a racionalidade do contrato, desde o ponto de vista individual, não deve ser confundida com a sua estabilidade. Mesmo modelando o “estado de natureza” como uma situação de interação entre indivíduos estrategicamente racionais,<sup>310</sup> surge um problema fundamental, que aponta para a necessidade do “contrato social”. Segundo a teoria dos jogos, existe, para cada um que escolhe, uma *estratégia*<sup>311</sup> *dominante*, isto é, que lhe garante o melhor resultado independentemente de qual seja a estratégia adotada pelos outros. O núcleo do problema é justamente o fato de que, uma vez adotada coletivamente, esta mesma estratégia produz um resultado estritamente inferior para todos. No apêndice II, procuro esclarecer este ponto, no qual repousa o interesse teórico do problema analítico constitutivo do jogo conhecido como o “dilema do prisioneiro”.

Quanto à estabilidade, a estratégia dominante é “um resultado que ambos os jogadores [podem] fazer acontecer e cuja melhoria não [pode] realisticamente ser melhorada do ponto de vista de cada um dos jogadores ... [É] conceitualmente análoga a um equilíbrio resultante de duas forças opostas”.<sup>312</sup> Este resultado é aquele para o qual

<sup>310</sup> Quer dizer, modelando-o como um jogo, operação que envolve os seguintes passos. Para um jogo de dois jogadores, podemos iniciar definindo as motivações e interesses básicos de *A*, o grau de informação de que cada um dispõe sobre *B* (seus interesses - que podem ser descritos pela “ordem de preferência dos resultados” -, que podem ser coincidentes ou conflitantes, total ou parcialmente, suas estratégias de escolha, o grau de informação de que dispõe sobre *A*). A seguir, supondo que os jogadores são racionais, a teoria tenta reconstituir o raciocínio levado a cabo por cada um deles e, a partir disso, encontrar a solução. No caso do contrato em questão, temos, na verdade, uma seqüência de um jogo que, a cada momento, se realiza simultaneamente entre múltiplos jogadores, cada um enfrentando um adversário de cada vez.

<sup>311</sup> Tecnicamente: a especificação de um programa completo contendo as regras de escolha para cada lance de um jogador numa “partida”.

<sup>312</sup> Anatol Rapoport, *Lutas, jogos e debates*, p. 105.

convergem várias deliberações individuais, e por isso se pode dizer que ele possui *estabilidade*. Que os jogadores jogam todos da melhor forma possível significa: eles preferem estratégias que otimizam os seus ganhos, e, ainda que sejam supostamente *egoístas racionais* - i. e., desejam e buscam o maior ganho possível *a longo prazo* - também estão supostamente *bem informados* sobre as condições que, no momento da escolha, determinam o que é melhor fazer. O cerne do problema é: o ponto de equilíbrio representado pela solução do “dilema do prisioneiro” não implica - e mesmo frustra - a realização do referido “equilíbrio a longo prazo”. Quer dizer, se mantidas as condições que definem a ordem de preferências individuais quanto aos tipos de ação a serem adotados por agentes racionais no “estado de natureza”, não se pode garantir, segundo a análise da teoria dos jogos, que aconteçam sucessivas interações com solução cooperativa entre um número cada vez maior de agentes, simplesmente porque em tais condições cooperar não é o melhor que cada um pode esperar fazer por si mesmo nos diferentes momentos em que interage com cada um dos demais.

A necessidade do contrato poderia então ser assim entendida: não é possível obter a chamada “estabilidade a longo prazo” se não for possível conseguir uma solução cooperativa entre aqueles que interagem na situação descrita como “estado de natureza”. Ou, a solução não-cooperativa é um meio incompatível com o interesse que define o fim principal de um *egoísta racional*. Vale lembrar que o fundamento do conflito no “estado de natureza” é a incompatibilidade existente entre a atitude de busca cega de segurança por meio da maximização da satisfação imediata dos próprios desejos (i. e., a atitude de um “egoísta crasso”) e a busca de segurança dirigida pela prudência racional, quer dizer, a busca da maximização dos ganhos definidos pelo interesse superior que um egoísta racional tem, i. e., o interesse no bem-estar da sua vida como um todo (ver 5.2 e nota 255).

As condições moldadas pelo soberano visam justamente a reconciliação de interesses cujo núcleo do conflito pode ser assim expresso: a busca individual guiada pelo princípio de maximização imediata do próprio bem-estar leva, quando generalizada, a resultados que não satisfazem as condições para a realização do fim mais desejado por cada um. A mais importante destas condições é a eliminação da dominância da solução composta pelas estratégias individuais não-cooperativas.<sup>313</sup> Assim, desde uma perspectiva

<sup>313</sup> Este é o ponto central do apêndice II.

coletiva, a escolha racional do soberano é, dado que há incerteza sobre o tipo de comportamento adotado pelos cidadãos, a permanente vigilância e a imposição de penas pesadas a todo aquele que violar as leis que direcionam “os esforços de cada pessoa, ocupada apenas com a busca do benefício próprio, ao ótimo social”.<sup>314</sup> Sob estas condições - em que deixa de ser racional, para cada indivíduo,<sup>315</sup> a busca cega do interesse próprio - está garantido o *equilíbrio a longo prazo*.

A estabilidade a longo prazo só pode ser assegurada pelo fato de um número cada vez maior de jogadores não poder se afastar da solução cooperativa sem deflagrar imediatamente acontecimentos que redundarão em desvantagem. Somente o retorno àquela solução restauraria as perdas ligadas ao que cada jogador supostamente mais valoriza, a saber, aquelas que ocasionam frustração do interesse que o define como egoísta racional.

Admitindo que a sociedade represente um estado de equilíbrio, segue-se, por definição, que em sociedade cada indivíduo não pode agir senão cooperando se quer fazer o melhor para si mesmo. Por isso mesmo, sair do *estado de natureza* envolve a adesão individual a normas sob condições tais que cada indivíduo pode confiar na permanente adesão dos demais<sup>316</sup> às mesmas normas; dito de outro modo, a sociedade é mantida por meio de normas que são *coletivamente estáveis*, mesmo sendo um estado de coisas individualmente inacessível e, às vezes, individualmente instável (quer dizer: algumas pessoas insistirão em considerar, pelo menos algumas vezes, mais vantajoso para si não cooperar).

<sup>314</sup> David Gauthier, *Morals by agreement*, p. 163.

<sup>315</sup> Sublinho a dependência de uma pressuposição adequada a respeito do “horizonte de tempo” que tais indivíduos levam em conta para computar os ganhos e perdas resultantes de suas decisões. A expressão “horizonte de tempo” foi tomada de empréstimo do professor João Carlos Brum Torres, quem a empregou no artigo *Discutindo a lei de Gêrson*; in Rohden (coord.) *Racionalidade e ação*, pp. 165-178.

<sup>316</sup> Possivelmente nem todos os demais.

## APÊNDICE I:

### Existe uma dualidade de critérios de valor na ética de Hobbes?

Por um lado, há a concepção hobbesiana de que dizer que uma coisa é útil é fazer um *juízo de valor* pelo qual cada um pode expressar sua expectativa de um prazer futuro, juízo que supõe a atribuição de poder a uma coisa, em que esta é concebida como meio cujo uso (emprego ou usufruto) é capaz de produzir prazer. Este seria um juízo de valor no qual se avalia, para usar um termo não-hobbesiano, o *valor de uso* dessa coisa.

Por outro, existe a concepção hobbesiana de acordo com a qual o valor de uma coisa é seu *valor de troca*<sup>317</sup>, isto é, seu *preço*. Neste último caso, poderíamos dizer que este valor expressa uma medida "julgada pelo mercado"; com isto significando que é um valor de algum modo resultante da interação das vontades de compradores e vendedores (o "apetite dos contratantes", na expressão de Hobbes) que buscam maximizar a satisfação de seu próprio interesse.

Esta maximização só é uma operação possível se cada um julgar o que quer segundo uma *escala individual de utilidade* - escala que varia conforme os efeitos (tipos de prazeres aos quais se associam, a cada momento do tempo, diferentes intensidades) de cada coisa no seu corpo. A variação destes efeitos é o que nos permite falar de diferentes graus de necessidade individual de um determinado bem.

Assim, a dualidade de critérios de valor em Hobbes poderia ser expressa da seguinte forma:

1. O *valor* de uma coisa qualquer é a sua *utilidade*, quer dizer, o quanto alguém a julga como capaz de produzir prazer;
2. O *valor* de uma coisa qualquer é seu *preço*, quer dizer, o quanto seria dado pelo uso de seu *poder*.

Estamos diante de duas proposições aparentemente incompatíveis. Se o valor de um bem é a medida da sua utilidade e esta medida está expressa no seu preço, então *utilidade* e *preço*, ainda que não se igualem em números absolutos, devem relacionar-se como grandezas diretamente proporcionais. A incompatibilidade em questão pode ser mostrada por meio do chamado "*paradoxo do valor*":<sup>318</sup> para uma pessoa que mora numa região árida, a água é um bem de grande utilidade - do seu uso diário, ainda que racionado, depende a sua atividade agrícola; a sua própria sobrevivência, afinal - pelo qual, ainda assim, pode não pagar um preço alto, se comparado ao preço dos diamantes, cujo uso, ao contrário, pouca satisfação lhe traria, dado o seu modo de viver. Este exemplo serve para

<sup>317</sup> Os termos "valor de uso" e "valor de troca" não são hobbesianos; foram usados primeiramente por Adam Smith.

<sup>318</sup> Uma versão clássica deste paradoxo tem o seguinte enunciado: "a água é muito útil e não tem valor, enquanto que os diamantes são inúteis, porém valiosos". D. P. O'Brien, *The classical economists*; p. 116 na ed. espanhola. As teorias do valor propostas pelos economistas clássicos visavam, entre outras coisas, oferecer uma solução para o "paradoxo do valor".

mostrar que não é necessário que *utilidade* e *preço* sejam grandezas diretamente proporcionais. Portanto, os enunciados 1 e 2 são incompatíveis: as generalizações por eles expressas não podem ser ambas verdadeiras.

Se, como diz Hobbes, o “valor de todas as coisas contratadas é medido pelo *apetite dos contratantes*” e também é *seu preço*<sup>319</sup>, uma pergunta segue-se quase naturalmente: de que modo os preços “refletem” tais apetites? Uma pista inicial para respondê-la é a afirmação de que, para todas as coisas<sup>320</sup>, é “o comprador quem determina o preço”.<sup>321</sup> Pesa então, na determinação do preço de um bem, a intensidade do desejo de quem o procura, o que o economista clássico M. Longfield definiu - com grande clareza, e pela primeira vez - como “intensidade da demanda”:

*“A demanda pode ser completamente atribuída à utilidade. A medida da intensidade da demanda por qualquer mercadoria de qualquer pessoa é a quantidade que estaria disposta e em condições de entregar para não renunciar a ela ou ver-se privada da gratificação que ... estima poder proporcionar dita mercadoria”*<sup>322</sup>

Mas que “quantidade” é essa a que Longfield se refere? Claramente não me interessa fazer uma análise do modo como os conceitos de “intensidade da demanda” e “utilidade” de fato se relacionam (e cada um deles, por sua vez, com o conceito de “preço”) na teoria desse economista. Interessa-me sim, fazendo uso destes e de outros conceitos, estabelecer uma justificação que seja, senão a mais adequada, pelo menos bem plausível para a seguinte proposição (que se segue das afirmações de Hobbes acima citadas):

*O preço é determinado pelo apetite dos compradores.*

<sup>319</sup> Cf *Leviathan*, 15, 75 e 10, 42 respectivamente; grifo meu.

<sup>320</sup> Hobbes tem em vista o que pode ser negociado (comprado e vendido) no mercado; tudo o que podemos chamar de *mercadoria* (“*commodity*”).

<sup>321</sup> *Leviathan*, 10, 42.

<sup>322</sup> *Lectures on political economy*; citado por D. P. O’Brien, *Op. cit.*, p. 151. A definição de “intensidade da demanda” deu origem à análise do conceito de *utilidade marginal* - que permite tomar o preço de uma mercadoria como função da utilidade, e não do *valor de uso* - tal como primeiramente desenvolvida pelos economistas neo-clássicos:

“Se [um indivíduo] é sensato ... persegue a máximo de sua utilidade total mediante uma igualação da utilidade marginal (o acréscimo de utilidade que se deve esperar de um pequeno acréscimo no volume comprado) de um dólar gasto em cada tipo de mercadoria e a utilidade que se espera da poupança de um dólar”. Joan Robinson e John Eatwell, *An introduction to modern economics*; p. 49 na edição mexicana.

Voltemos agora ao enunciado 2. Nele, afirma-se que o preço de uma coisa qualquer indica o “quanto seria dado pelo uso de seu poder”<sup>323</sup>. Mas, perguntemos, “dado por quem”? Em princípio, por cada um dos compradores, porque a intensidade do desejo de cada um determina (pelo menos em parte, segundo a teoria microeconômica moderna) a demanda individual por um bem ou serviço (uma mercadoria). O fato de alguém estar disposto a pagar mais ou menos por um produto qualquer depende - pensemos especificamente no caso dos bens essenciais para a vida<sup>324</sup> - da intensidade do seu desejo por tal produto. Mas, poderíamos acrescentar, também depende da intensidade do desejo de outros indivíduos pelo mesmo produto. Em última análise, o preço que os indivíduos pagam por uma mercadoria, num determinado momento, depende do apetite de todos os seus compradores potenciais.<sup>325</sup>

Mesmo que seja desconsiderada a explicação sobre o modo como apetites individuais de diferentes intensidades são “agregados” de forma a determinar o *valor econômico* de uma coisa - isto é, o *preço* a que ela é comprada e vendida - , uma coisa é falar do valor que alguém termina de fato pagando por ela e outra, do quanto estaria disposto, hipoteticamente falando, a pagar pelo uso do seu poder. O valor econômico de um produto não é necessariamente uma medida da utilidade que uma pessoa lhe atribui, pois o preço que os indivíduos pagam por ele é, por assim dizer, uma medida da “utilidade agregada” do seu mercado consumidor.

Assim, independentemente da avaliação que fizemos da validade da *teoria do valor* de Hobbes como parte de uma teoria econômica, pode-se dizer que, do ponto de vista conceitual, a dualidade de critérios de valor naquela *teoria* é apenas *aparente*.

<sup>323</sup> Grifo meu. O emprego do condicional é um aspecto essencial para o conceito de preço que está sendo usado: o valor de todas as coisas, afirma Hobbes, é seu preço, quer dizer, “so much as would be given for the use of his Power”. *Leviathan*, 10, 42. O condicional que define o preço de uma coisa é *contrafático*: “se um comprador pudesse determinar isoladamente o preço de uma coisa *x* qualquer, então tal preço equivaleria ao valor atribuído pelo comprador ao seu próprio uso do poder de *x*”.

Hobbes está esboçando uma resposta ao problema que só ganhará um tratamento analítico-quantitativo preciso por meio dos chamados “mapas de indiferença”: a relação conceitual entre *utilidade* e *preço*. Tais mapas foram idealizados por Edgeworth e analisados na obra *Mathematical psychics: an essay on the application of mathematics to the moral sciences*, publicada em 1881. Uma apresentação introdutória dos “mapas de indiferença” é feita por Charles Dyke em *Philosophy of economics*; pp. 67-73 na trad. argentina.

<sup>324</sup> Por exemplo, os alimentos que contêm aminoácidos que não podem ser sintetizados pelo corpo humano. O seu consumo, contudo, não é visto como igualmente necessário pelos diferentes consumidores. Estou omitindo uma outra complicação: a existência dos chamados *bens substitutos*, bens que se relacionam de tal modo que uma variação no preço de um afeta a procura do outro.

<sup>325</sup> Há aqui um pressuposto forte: que a quantidade oferecida de um produto não afeta os preços; quer dizer, que não há problemas de *escassez*.

## APÊNDICE II:

### Como é possível escapar do estado de natureza?

O problema fundamental a ser resolvido pela *moralidade*, tal como Hobbes a concebe, pode ser assim enunciado: dada a predominância de motivações egoístas entre homens igualmente situados (i. e., iguais quanto à força e capacidades mentais, e à deficiência de informação sobre os outros) e - mesmo na existência de um grau reduzido de *benevolência* - sem qualquer garantia da realização de ações cooperativas, como é possível eliminar a dominância de um comportamento cujas conseqüências são um conflito generalizado e uma radical insegurança na manutenção da vida; sumariamente, uma ameaça permanente à realização do que cada um mais deseja, a *própria felicidade*?

E, como solução deste problema - cujo núcleo é de caráter motivacional - não basta indicar quais regras são capazes de gerar - se seguidas coletivamente - um padrão de comportamento que seja suficiente para contrabalançar até mesmo uma completa ausência de *benevolência*.<sup>326</sup>

Como vimos, supõe-se que as forças motivacionais determinantes na psicologia da pessoa moral (ver 1.2) hobbesiana derivam da chamada "disposição para a paz" (ver 5.11). Esta disposição é suficiente, quando satisfeitas algumas condições (ver 5.15), para justificar a adesão da pessoa moral hobbesiana às "leis da natureza": cada pessoa reconhece que é racional buscar a paz porque esta é o meio mais seguro para garantir a continuidade da satisfação dos seus próprios desejos (ver 5.14). É suficiente tomar a paz como meio para a felicidade individual: não é necessário supor um acordo universal sobre o modo de dar conteúdo ao bem individual.

Um agente é racional sempre que age de modo a maximizar seus ganhos (dito desde outro ângulo: sempre que age buscando minimizar suas perdas). É possível esquematizar o princípio que guia a deliberação e a escolha de um agente racional. Este princípio condensa a concepção de racionalidade como *maximização*:

Se *A* é racional, então se

- (i) *A* sabe que pode fazer *x* ou *y*, mas não pode fazer ambas; e
- (ii) *A* sabe que se faz *x*, ganha  $g_1$  (não importa qual a medida adotada para computar este ganho), e se faz *y*, ganha  $g_2$ ; e

<sup>326</sup> Se há algum motivo racional para agir moralmente, ele não pode se resumir a seguir regras cuja rigidez e incondicionalidade estão além do que pode ser racionalmente ponderado. Quer dizer, não se pode agir moralmente se não se pode fazê-lo deliberadamente. Isto implica que no pensamento capaz de justificar a conduta moral (independentemente do fato de ter ou não de fato antecedido a ação) inclui-se, necessariamente, a operação de *pesar razões*.

Pesar razões é uma operação constituinte da *deliberação* (ver 4.2). Na teoria hobbesiana, as razões são sempre "*relativas a desejos*" (ver 4.7). São nestes desejos que estão baseados os diferentes motivos ou razões para agir (ver 1.3 e 1.5); dizendo-o por meio de uma imagem, pesar razões equivale a calcular a resultante de um jogo de forças (supondo que os desejos podem ser, sem maiores problemas, assim representados).



- (iii)  $A$  sabe que  $g_1 > g_2$ ,  
então  $A$  deseja fazer  $x$  (em vez de  $y$ )<sup>327</sup>.

Ganhos e perdas são medidos conforme *escalas individuais de valores*. Uma maneira de comensurar valores é atribuir às conseqüências das ações um valor numérico; esta atribuição é o principal papel das chamadas *funções de utilidade* (ver 4.6). É um pressuposto forte da teoria do valor de Hobbes a existência de uma *função de utilidade interpessoal*, a que atribui valor àquelas ações que contribuem para a paz, independentemente da escala que cada indivíduo usa para medir seus próprios ganhos e perdas. O acordo expresso por tal função de utilidade diz respeito ao valor das condições sob as quais cada indivíduo pode razoavelmente esperar atingir seus próprios fins ou objetivos.

A existência desta função tem uma implicação muito importante. Para qualquer indivíduo, todo e qualquer estado social - no qual ele ocupa uma determinada posição, à qual estão associadas determinadas ações e respectivas conseqüências em termos de bem-estar pessoal - é preferível ao estado de natureza, pois, independentemente da posição social deste indivíduo e do modo como ele avalia as conseqüências a ela associadas, ele sempre está numa situação que julga melhor que qualquer outra em que estaria caso permanecesse naquele "estado de guerra".<sup>328</sup>

O termo "solução" aplica-se a uma das duas, ou ambas, respostas às seguintes perguntas: (a) "como devem comportar-se os jogadores?" e (b) "qual deve ser o resultado

<sup>327</sup> Se o desejo de fazer  $x$  é "moldado" pelas condições expostas em *i*, *ii* e *iii*, podemos denominá-lo de "desejo racional" de fazer  $x$ . Outras considerações epistemológicas podem informar a cadeia deliberativa, resultando, ao contrário, na formação do desejo de fazer  $y$ . Na concepção hobbesiana, pode-se dizer que  $A$  é livre para fazer  $x$  se  $A$  desejar fazer  $x$  e nenhum obstáculo externo ~~impede~~<sup>impedir</sup>  $A$  de fazê-lo. O termo *liberdade* designa para Hobbes a ausência de interferência de obstáculos externos às ações dos indivíduos que buscam satisfazer seus desejos. Em sociedade, há uma limitação legítima desta liberdade: nem tudo o que cada um quer fazer deve continuar livre de obstáculos, pois toda ação que representar uma ameaça à preservação da vida de qualquer um dos outros não pode ser racional nem mesmo para aquele que quer realizá-la. Ao poder soberano cabe a tarefa, por meio do estabelecimento de sanções adequadas a cada caso, de torná-la irracional.

<sup>328</sup> Podemos comparar e hierarquizar o bem-estar interpessoal comparando de modo apenas ordinal o bem-estar relativo a uma mesma posição social em diferentes estados sociais. Por exemplo, "É melhor ser o senhor  $A$ , um trabalhador cronicamente desempregado, numa sociedade cujo governo tem uma maior ou menor tendência assistencialista?" Ou ainda, comparar diferentes posições em diferentes estados sociais. Por exemplo: "É melhor ser o senhor  $A$ , um trabalhador não-especializado desempregado, num país sem recessão, mas com economia estagnada; ou ser o senhor  $B$ , um engenheiro especializado, num país com estagnação crônica do setor da produção industrial no qual se especializou?". Mas, a comparação interpessoal assumida por Hobbes poderia ter o seguinte sentido: "Mesmo que você tome o pior estado social que possa conceber, digamos nós - para dar um exemplo contemporâneo- uma ditadura militar de direita num país sul-americano em plena guerra fria, e que se imagine na pior posição neste estado, qual seja, a de um militante de esquerda que é obrigado a viver num completo anonimato, por estar praticamente privado de suas liberdades políticas (de opinião, de reunião, de associação, etc.), você ainda assim não estaria vivendo melhor do que numa situação em que organizadamente participasse de uma guerra civil aberta e incessante contra esse governo?"

final do jogo?". O termo "deve" não tem um sentido moral, mas significa aquele comportamento que leva o jogador ao *melhor resultado* - que *deve* ser o final - supondo os seus interesses (hierarquizados por meio das "preferências" do jogador pelos diferentes resultados), a regra de escolha racional que seguem, o fato de suas escolhas serem independentes ou não (isto é, se a escolha que cada um faz influi ou não na escolha do outro e/ou no resultado) e se sabem ou não qual é o caso, o tipo de informação de que dispõem sobre a situação de escolha (por exemplo, o quanto conhecem dos "hábitos" de cada adversário, isto é, se podem prever ou não as chances de escolha de cada uma das estratégias possivelmente adotadas) e o grau de comunicação possível entre eles (e o que pode fundamentalmente decorrer disso, como por exemplo, que eles façam acordos).

Suponhamos um agente *B*, também racional. Suponhamos um tipo de situação interativa entre *A* e *B* estruturada de tal modo que os ganhos de cada um são resultados que dependem também da escolha feita pelo outro. A principal dificuldade, para cada um deles, consiste em *coordenar* a sua escolha com as escolhas possíveis do outro de modo a - uma vez que guiam suas escolhas pelo princípio de *maximização* dos ganhos - conseguir obter, para si, o maior ganho possível; dito de outra forma, conseguir fazer uma *escolha ótima*.

Podemos representar, por meio de uma matriz de ganhos (que define a estrutura de uma *situação interativa* ou *jogo* em questão), as escolhas possíveis para cada um dos agentes e os resultados a elas associados. O que define a *estrutura de um jogo* são as relações que se dão entre os diferentes resultados. Diferentes conjuntos de valores numéricos para os quais são válidas as mesmas relações definem *variantes* de um mesmo jogo. As relações relevantes para uma teoria da *escolha racional* (descritiva e normativa) são aquelas que correspondem a uma hierarquia de preferência entre as alternativas de ação que conduzem a cada resultado.

Assim, definimos o jogo denominado como "o dilema do prisioneiro" (referido de agora em diante como "*DP*") pelo fato de que, para cada agente, as alternativas de escolha individual (representadas por *C* e *D* na matriz de ganhos associados às *ações combinadas* ou, usando um termo de uso mais freqüente, *coletivas*) são ordenadas conforme uma determinada ordem de preferência:

		agente <i>B</i>	
		C	D
agente <i>A</i>	C	x	y
	D	z	w

\*Uma representação mais comumente encontrada desta matriz inclui valores numéricos relativos dos ganhos de cada jogador - o número no canto inferior esquerdo de cada quadrante indica o ganho de *A*; aquele no canto superior direito, o ganho de *B*:

	1	2
1	-1	
2	-1	0

Neste caso, a ordem de preferência individual das ações é:

- (1) {D, C}
- (2) {C, C}
- (3) {D, D}
- (4) {C, D}

As ações *C* e *D* são interpretadas, em cada variante, como tipos de *ações cooperativas* e *não-cooperativas* respectivamente. Além disso, seguindo uma sugestão de Elster, podemos fazer o *DP* aparecer como um jogo entre “eu” (cujas alternativas de escolha estão representadas pela 1<sup>a</sup> variável do par ordenado das preferências) e qualquer um dos demais que pode interagir comigo (as escolhas de qualquer um de “todos os outros” estão sempre representadas pela 2<sup>a</sup> variável). Como as ações *C* e *D* são escolhidas para a interação com todos os possíveis “jogadores” com que “eu” me defronto, o que propriamente escolho é uma *estratégia* de ação.<sup>329</sup>

A estrutura da *matriz 2x2* acima e a estrutura da ordem individual de preferência ajudam-nos a explicitar o tipo de raciocínio levado a cabo por cada jogador nesta situação de escolha: “se um qualquer dos demais coopera, prefiro não cooperar; se não coopera, também prefiro não cooperar; logo, qualquer que seja a sua escolha, o melhor que posso fazer por mim mesmo é não cooperar”. “Não cooperar” é portanto, neste tipo de situação, a estratégia racional para mim” - raciocina cada jogador.

Todavia, se, a cada interação, ambos jogadores raciocinam da mesma maneira, a ação combinada é {D, D}. Mas esta não produz o melhor resultado para *cada um* deles, pois *ambos* estariam melhor na situação resultante da ação {C, C}. E isto acontece como resultado de um raciocínio estratégico, no qual cada um leva em conta que o outro também busca o melhor para si mesmo e que suas próprias expectativas dependem (como uma *função*) das expectativas do outro sobre elas.<sup>330</sup>

O resultado *coletivamente ótimo* - os ganhos representados pela situação referida por “*x*” - pode ser observado na própria ordem dos ganhos para cada agente (“>” está por “... é maior que ...”):

para *A*:  $z > x > w > y$

para *B*:  $y > x > w > z$

O par de estratégias {D, D} é considerado um *ponto de equilíbrio*, isto é, *D* é a melhor resposta de *A* se *B* escolhe *D*; valendo também a proposição inversa. Este par é também denominado de *solução não-cooperativa*, quer dizer, não é um resultado alcançado graças a algum tipo de compromisso (acordo, promessa ou mesmo ameaça).

<sup>329</sup> Para a definição de *estratégia* na *teoria dos jogos*, ver nota 311.

<sup>330</sup> Se estendemos o *DP* de duas pessoas para grupos maiores, a expressão “melhor para todos” deixa de ser equivalente à generalização distributiva “melhor para cada um”. Isto porque “outro” significa “todos os outros” quando há somente dois jogadores, mas admite possibilidades intermediárias (“nem todos os outros”) nas versões para mais de dois.

Mesmo supondo a existência de algum tipo de comunicação entre os jogadores, tornando possível um compromisso explícito, sem ter como *impor* ao outro a obediência ao compromisso, como estar seguro da disposição dele a fazer o que se compromete a fazer? Dadas as condições de escolha no estado de natureza hobbesiano, ninguém pode sentir-se razoavelmente seguro a respeito da adoção de uma ação cooperativa por parte dos demais.

Considerações sobre a *simetria situacional* dos jogadores só tendem a reforçar a adoção da solução não-cooperativa. Vejamos um exemplo: mesmo supondo - já que comumente conheço mais a respeito de "minhas próprias" intenções - que os outros jogadores são tão racionais e astutos quanto "eu mesmo", por que, esperando que eu cooperasse, não deixariam de cooperar, do mesmo modo que "eu" faria se esperasse a cooperação por parte deles? O ponto é que, devido à ordem de preferências definidora do *DP*, não parece haver como se produzir outro resultado.<sup>331</sup>

Podemos sublinhar alguns resultados, dizendo que o *DP* mostra que:

i) a combinação de ações racionais individuais podem produzir um resultado sub-ótimo (dado o conjunto de ações combinadas possíveis, existe uma outra ação cujas conseqüências seriam melhores para ambos indivíduos);

ii) o resultado, no entanto, representa o melhor que cada jogador pode escolher para si, dada a escolha provável do outro jogador (tecnicamente, a solução está em "equilíbrio"; para a interação estratégica em questão, existe o chamado "equilíbrio em estratégias dominantes").

O item *i* revela um ponto importante: uma interação estratégica não é uma condição suficiente para a chamada "racionalidade coletiva" (ver 5.11). O item *ii* explicita a seguinte dificuldade: estamos diante de uma situação *estável*, quer dizer, se cada um chega à conclusão que qualquer um dos demais escolherá a ação não-cooperativa, cada um deixará de realizar, em primeiro lugar, a ação cooperativa; ao mesmo tempo, se qualquer um dos demais porventura realiza primeiro a ação cooperativa, também é racional (dada a ordem de preferências assumida) para cada um deixar de cooperar.

De um exame mais acurado destes pontos depende a plausibilidade da afirmação hobbesiana de que é possível escapar do "estado de natureza". Isto porque, se aceitamos que este estado é uma instanciação da estrutura geral expressa no *DP*, como é possível obter a *solução ótima*? (Lembrando que esta solução significa: uma *ação coletiva* - porque, num jogo, a solução é sempre uma *ação combinada* - cuja conseqüência é a melhor possível para os indivíduos nela envolvidos.)

Uma maneira de traduzir neste caso a pergunta "como obter a solução ótima?" seria "como fazer com que, na situação do *DP*, as pessoas deixem de praticar a ação

<sup>331</sup> As tentativas de argumentação a favor da cooperação no *DP* podem consistir em procurar minar alguma premissa do chamado "argumento da dominação", i. e., o argumento por meio do qual se sustenta que a solução racional do *DP* é *sempre* aquela representada pela estratégia dominante. Randall Campbell (ver bibliografia) analisou uma destas tentativas, o chamado "*argumento da simetria* a favor da cooperação". Este autor conclui que, ao contrário do que parece à primeira vista, tal argumento é falacioso porque simplesmente usa como premissa a negação da conclusão do "argumento da dominação", sem oferecer, no entanto, nenhum argumento independente a favor dessa premissa.

dominante, aquela que cada uma pensa ser racionalmente do seu próprio interesse praticar?”. Primeiramente, trata-se de assumir a matriz de ganhos e perdas como expressando corretamente a ordem de preferência de cada indivíduo na situação interativa descrita pelo *DP*. Esta ordem deve ser a que melhor expressa a estrutura psicológica do agente hobbesiano assim situado. E, fazendo esta pressuposição, perguntar - tal como sugerido por John Harsanyi<sup>332</sup> - pela “estrutura de incentivos” capaz de alterar a solução do *DP*, fazendo que o par de estratégias {C, C} passe a representar a ação combinada que os jogadores podem *adotar* racionalmente; podendo cada um também *esperar* racionalmente tal adoção por parte do outro.<sup>333</sup>

Existem duas perspectivas pelas quais podemos ver melhor por que afinal o núcleo do problema que estamos analisando neste apêndice é, como dissemos logo no início, de caráter motivacional. A primeira diz respeito ao reconhecimento da estrutura da matriz de ganhos do *DP* como expressando de fato os traços definidores da situação descrita por Hobbes como “estado de natureza”. A segunda, de como é possível resolver o problema de *ação coletiva* com o qual nos deparamos uma vez que damos por assentada a primeira perspectiva.

Pressupondo a adequação do *DP* à situação em questão, há uma consequência importante, pois, se é verdadeira, para a teoria ética de Hobbes, a afirmação de que “o dilema do prisioneiro é a representação formal do tipo de interação estratégica que caracteriza o estado de natureza”, então *não* se segue que “a motivação que atravessa os comportamentos no estado de natureza [é] o egoísmo defensivo ... que conduz os dois cooperadores condicionais, cada um dos quais preferiria cooperar desde que o outro cooperasse, à violação de cooperar”.<sup>334</sup> *Não se segue* porque a estrutura do *DP* não pode expressar as preferências de *cooperadores condicionais*, independentemente de quais sejam de fato as suas escolhas ou omissões.

Cooperadores condicionais têm suas preferências expressas pelo chamado “*assurance game*”. Amartya Sen foi quem originalmente introduziu este jogo, cuja estrutura foi construída a partir do *DP*. A ordem de preferência individual deste jogo é:

- (1) {C, C}
- (2) {D, C}
- (3) {D, D}
- (4) {C, D}

A ordem dos ganhos é:

para A:  $x > z > w > y$

para B:  $x > y > w > z$

<sup>332</sup> *Advances in understanding rational behavior*, in Ester (ed.), *Rational choice*, p. 95.

<sup>333</sup> Cf. a definição de “solução” apresentada por Harsanyi, *Op. cit.*, p. 94.

<sup>334</sup> Tavares, José Antônio G., *Dois paradigmas de ética pública na teoria liberal*, pp. 348 e 351. Tirar tal consequência é um dos erros localizados deste artigo abrangente e esclarecedor.

O “estado de natureza” hobbesiano é uma situação que pode ser caracterizada como uma extensão que generaliza o DP a partir de duas pessoas. Aceita tal extensão, temos um grupo que enfrenta um *problema de ação coletiva*. Este problema pode ser definido por meio das preferências que guiam o raciocínio que cada indivíduo racional realiza a fim de decidir o que fazer nas suas interações com os demais. No “estado de natureza” cada um prefere que os demais cooperem a que não cooperem (é melhor, para cada um, que pelo menos alguns dos demais cooperem do que ninguém o faça) e prefere não cooperar cada vez que os demais cooperem (é melhor, para cada um, deixar de cooperar quando pelo menos alguns dos demais cooperam). Estas preferências geram uma tendência ao “desastre coletivo”, o que não pode ser evitado, já que neste caso a cooperação não pode emergir sem algum mecanismo de coordenação de interesses centralizado. Não pode aí existir coordenação descentralizada e não-compulsória porque as ações não-cooperativas são estrategicamente dominantes.

Por princípio, não se pode alterar a psicologia moral do agente hobbesiano. Desta forma, não basta, para resolver um problema de ação coletiva,<sup>335</sup> buscar persuadir tais agentes que devem agir *como se* seus acordos pudessem ser cumpridos não compulsoriamente ou que devem confiar nas outras pessoas, *como se* tivessem boas razões para fazê-lo. Apelar para este tipo de *dever* - esperando que por si só possa determinar a motivação individual - significa supor erroneamente que o agente hobbesiano possa motivar-se por resultados sob circunstâncias *hipotéticas* e não somente por resultados efetivos de suas ações sob circunstâncias também efetivas.

A solução para o problema depende então da mudança dos prêmios associados a cada ação combinada possível de forma a eliminar a dominância da ação {D, D}. É preciso mexer na estrutura de incentivos externos, por meio da imposição de sanções, de forma que *cooperar* represente o melhor que cada um pode fazer por si mesmo. Como vimos, o termo “melhor (o que é racional)” é relativo aos interesses de cada agente, hierarquizados segundo uma ordem. Dada a ordem de preferência que define um agente como *egoísta racional* (ver 4.13 e 5.2), o “melhor que pode fazer por si mesmo” é a busca deliberada e racional da satisfação do seu auto-interesse de longo prazo. Esta busca é guiada pelo *desejo de maximizar globalmente* (ver 2.9, 4.2, 4.10 e 5.2).

A imposição de sanções não contraria o interesse de um egoísta racional, apenas cria as condições - por meio de algum mecanismo de coordenação centralizada<sup>336</sup> - que tornam racional a sua cooperação. Runciman e Sen ajudam-nos a clarificar este ponto:

<sup>335</sup> Uma “solução” é um resultado que coordene os interesses de agentes racionais *qua* racionais, quer dizer, um resultado que represente, para cada um destes agentes, o maior ganho possível de ser obtido, diante das possibilidades de escolha suas e dos outros. É este o sentido da expressão “ótimo de Pareto”.

<sup>336</sup> Determinar em que tipo de instituições - pode-se definir uma instituição como um “mecanismo de imposição de regras” - este mecanismo deve se apoiar é um problema a ser solucionado pela *filosofia política* de Hobbes. A definição de *instituição* é de Jon Elster, *Nuts and bolts for the social sciences*; p. 174 na ed. brasileira.

*"[A imposição da racionalidade coletiva] não requer que nenhuma pessoa deva ser (em termos de sua própria ordem de preferência) um perdedor de longo-prazo [long-term loser], embora possa, é claro, exigir-lhe que adie a busca de uma vantagem individual que, sem a imposição de um conluio [enforced collusion], deixá-la-ia no final em pior situação."*<sup>337</sup>

---

<sup>337</sup> *Games, justice and the general will*, p. 557. O termo "collusion" tem um correspondente exato em português. "Colusão" significa, conforme o dicionário Aurélio, um "ajuste secreto e fraudulento entre duas ou mais partes, com prejuízo para terceiros".

### Apêndice III:

#### Uma justificação matemática elementar para a afirmação:

A autopreservação é alcançada por meio de um mecanismo de ascensão gradual e contínua da capacidade do corpo do homem de manter seu movimento vital; um tal mecanismo só pode operar por meio de transições contínuas - e indiretas - entre *máximos locais*, até alcançar um máximo local que coincida com o global (p. 46).

Seja  $f(t)$  uma função definida e contínua no intervalo aberto de tempo  $I(t_1, t_n)$  (i. e., um certo período de duração do movimento vital)

[Esta condição é garantida porque  $f(t)$  assume valores que indicam o *quantum* de prazer, cuja continuidade mesma define a duração da vida.]

*Definição 1:*  $f$  tem um máximo local em um ponto  $t_2$  pertencente ao intervalo aberto  $(t_1, t_n)$  se e somente se, para todo  $t$  pertencente a este intervalo,  $f(t_2) > f(t)$  ou  $f(t_2) = f(t)$ .

*Definição 2:*  $f$  tem um mínimo local em um ponto  $t_2$  pertencente ao intervalo aberto  $(t_1, t_n)$  se e somente se, para todo  $t$  pertencente a este intervalo,  $f(t_2) < f(t)$  ou  $f(t_2) = f(t)$ .

*Teorema 1* [Enuncia uma *condição necessária para extremos locais*]:<sup>338</sup>

Se  $f$  possui um *extremo local* em  $t_2$  (i.e., um máximo ou um mínimo locais), e é diferenciável neste ponto, então a derivada de  $f$  em  $t_2$  é igual a zero [ $f'(t_2) = 0$ ]

*Teorema 2* [*Teorema do valor médio* aplicado a um intervalo cujos extremos são máximos locais].<sup>339</sup>

Seja  $f$  uma função tal que:

- (i) é definida e contínua no intervalo fechado  $[t_1, t_n]$ ;
- (ii)  $f(t_1)$  e  $f(t_n)$  são diferentes;
- (iii)  $f$  é diferenciável no intervalo aberto  $(t_1, t_n)$ ;
- (iv)  $f(t_1)$  e  $f(t_n)$  são máximos locais.

<sup>338</sup> A prova deste teorema e do seguinte podem ser encontradas em qualquer manual de introdução ao Cálculo Infinitesimal. Cito os usados para desenvolver este apêndice: Munem, M. A. & Foulis, D. J. *Cálculo*. Editora Guanabara, 1982, vol. 1; Leithold, L. *O cálculo com geometria analítica*. Harbra, 1977, vol. 1; e Spivak, M. *Cálculo infinitesimal*. Editorial Reverté, 1970, vol. 1.

<sup>339</sup> Na verdade, este teorema agrupa dois teoremas (apresentados nos manuais como “teorema do valor intermediário” e “teorema do valor médio”) e está aqui aplicado ao caso relevante para a justificação ora em questão.



Então,

- a) existe pelo menos um valor  $t_2$  no intervalo aberto  $(t_1, t_n)$  tal que  $f(t_2)=k$  tal que  $k=f(t_1)/f(t_n)$ ;
- b) existe pelo menos um valor para  $t$  tal que  $f'(t)=0$ .

Dito de outra forma:

a) dados dois máximos locais como extremos do intervalo fechado  $[t_1, t_n]$  para o qual se define a função contínua “*quantum* de prazer”, ela assumirá todos os valores possíveis entre esses extremos, o que garante a existência do valor médio;

b) dados dois máximos locais como extremos do intervalo fechado  $[t_1, t_n]$  para o qual se define a função contínua e diferenciável “*quantum* de prazer” - com  $f'(t_1)=f'(t_n)=0$ , existe pelo menos um valor para a variável  $t$  tal que  $f'(t)=0$ , isto é, existe, entre esses extremos do intervalo, pelo menos um outro *extremo local*, o que significa que a passagem entre máximos locais não se pode dar *discretamente* (*descontinuamente* ou *por saltos*), mas apenas por uma ascensão gradual e, principalmente, que envolve intervalos de tempo para os quais a função é decrescente.

## BIBLIOGRAFIA

- AGOSTINHO, Santo. De magistro. Abril Cultural, 1983.
- ALLISON, Henry. Morality and freedom: Kant's reciprocity thesis. The Philosophical Review, vol. 95, n° 3, july 1986, pp. 393-425.
- ANSCOMBE, E. Modern moral philosophy. Philosophy, vol. 33, n° 124, january 1958, pp. 1-19.
- AUSTIN, J.L. How to do things with words. Trad. brasileira: Quando dizer é fazer. Editora Artes Médicas, 1990.
- ARISTÓTELES. Éthique à Nicomaque. Vrin, 1987.
- BACON, Francis. Novum organum. Abril Cultural, 1984.
- BAIER, Kurt. Egoism in SINGER, P. (ed.) A companion to ethics.
- BARBOSA FILHO, Balthazar. Condições de autoridade e autorização em Hobbes. Filosofia Política 6, L & PM Editores, 1991.
- BENNETT, Jonathan. A study of Spinoza's ethics. Trad. mexicana: Fondo de Cultura Económica, 1990.
- BENTHAM, Jeremy. Uma introdução aos princípios da moral e da legislação. Abril Cultural, 1979.
- BERLIN, Isaiah. The crooked timber of humanity. Trad. brasileira: Limites da utopia, Cia. das Letras, 1991.
- Four essays on liberty. Trad. brasileira: Editora UNB, 1981.
- BITTNER, Rüdiger. What reason demands. Cambridge U. P., 1989.
- BLOOM, Allan. The closing of the american mind. Trad. brasileira: O declínio da cultura ocidental, Ed. Best Seller, 1989.
- CAMPBELL, Randall K. The prisoner's dilemma and the symmetry argument for cooperation. Analysis, vol. 49, 1989, pp. 60-65.
- CAMPBELL, Tom. Seven theories of human society. Trad. espanhola: Ediciones Cátedra, 1992.
- COPI, Irving. Introduction to logic. Trad. brasileira: Editora Mestre Jou, 1981.
- COTTINGHAM, John. Descartes. Trad. portuguesa: A filosofia de Descartes, Edições 70, 1989.
- CUBITT, Robin P. & HOLLIS, Martin. The mutual investment game: peculiarities of indifference. Analysis, vol. 51, n° 3, june 1991, pp. 113-119.
- DA FONSECA, Eduardo Giannetti. Vícios privados, benefícios públicos? Companhia das Letras, 1994.

- Beliefs in action: economic philosophy and social change. Cambridge U. P., 1991.
- DANIELS, Norman (ed.). Reading Rawls. Stanford U. P. 1989.
- DARWALL, Stephen. Motive and obligation in Hume's ethics. *Noûs*, vol. 27, n° 4, 1993, pp. 415-448.
- DAVIDSON, Donald. Las paradojas de la irracionalidad. *Análisis Filosófico*, vol. 1, n° 2, 1981, pp. 1-18.
- Rational animals. *Dialectica*, vol. 36, n° 4, 1982, pp. 317-327.
- DAVIS, M. Game theory: a nontechnical introduction. Basic Books, 1970.
- DEANE, Phyllis. The evolution of economic ideas. Trad. brasileira: Zahar, 1980.
- DESCARTES, René. Meditações & As paixões da alma. Abril Cultural, 1973.
- DWORKIN, Ronald. Foundations of liberal equality. Trad. espanhola: Ética privada e igualitarismo político, Ediciones Paidós, 1993.
- DYKE, Charles E. Philosophy of economics. Trad. argentina: Editorial Paidós, 1983.
- ELSTER, Jon. Nuts and bolts for the social sciences. Cambridge U. P., 1989. Trad. brasileira: Ed. Relume-Dumará, 1994.
- Ulysses and the sirens: studies in rationality and irrationality. Cambridge U. P. & Editions de la Maison des Sciences de L'homme, 1990.
- (ed.) Rational choice. New York U. P., 1986.
- ENGEL, S. Morris. Hobbes's "table of absurdity". *The Philosophical Review*, vol. 70, n° 4, 1961, pp. 533-543.
- FOLLESDAL, Dagfinn. The status of rationality in interpretation and in the explanation of action. *Dialectica*, vol. 36, n° 4, 1982, pp. 301-316.
- FRANKENA, William. Ethics. Trad. brasileira: Zahar, 1981.
- GAUS, Gerald. Value and justification. Cambridge U. P., 1990.
- GAUTHIER, David. Moral dealing. Cornell U. P., 1990.
- Morals by agreement. Oxford U. P., 1986.
- Morality and advantage. In RAZ (ed.) Practical reasoning.
- GAUTHIER, D. & SUGDEN, R. (eds.) Rationality, justice and contractarianism. The University of Michigan Press, 1993.
- GOLDMAN, Alan. Moral knowledge. Routledge, 1988.
- GOOD, J. Probability and the weighing of evidence. Charles Griffin, 1950.
- GOULD, Stephen Jay. Ever since Darwin. Trad. brasileira: Darwin e os grandes enigmas da vida. Martins Fontes, 1992.
- Hen's teeth and horse's toes. Trad. brasileira (parcial): Os dentes da galinha. Paz e Terra, 1996.

- GRICE, G. R. Motive and reason. In RAZ (ed.) Practical reasoning.
- HACKING, Ian. The emergence of probability. Cambridge U. P., 1975.
- HAMPSHIRE, Stuart. Morality and conflict. Harvard U. P., 1984.
- HAMPTON, Jean. Hobbes and the social contract tradition. Cambridge U. P., 1986.
- HARE, R. M. Freedom and reason. Oxford U. P., 1963.
- HARSANYI, John. Morality and the theory of rational behaviour in SEN, A. & WILLIAMS, B. (eds.) Utilitarianism and beyond.  
Advances in understanding rational behaviour. In ELSTER (ed.) Rational choice.
- HAWTHORN, G. (ed.) The standard of living. Cambridge U. P., 1987.
- HEAP, S. H. et al. The theory of choice: a critical guide. Blackwell, 1992.
- HIRSCHMAN, Albert. The passions and the interests. Trad. brasileira: Ed. Paz e Terra, 1979.
- HOBBS, Thomas. De cive ou les fondements de la politique. Éditions Sirey, 1981. Trad. brasileira: Martins Fontes, 1992. Tradução, apresentação e notas de Renato Janine Ribeiro.  
De homine. Trad. portuguesa: A natureza humana. Imprensa Nacional-Casa da Moeda, 1983.  
Leviathan. Penguin Books, 1968. Trad. brasileira: Abril Cultural, 1974.  
A dialogue between a philosopher and a student of the Common Law of England & Thomae Hobbes Malmesburiensis vita carmine expressa auctore ipso. Trad. espanhola: Diálogo entre un filósofo y un jurista y escritos autobiográficos. Tecnos, 1992. Estudo preliminar, tradução e notas de Miguel Ángel Rodilla.
- HOLLIS, Martin. Models of man. Cambridge U. P., 1977.  
Moves and motives in the games we play. *Analysis*, vol. 50, n° 2, march 1990, pp. 49-62.
- HONDERICH, Ted & BURNYEAT, Miles (eds.). Philosophy as it is. Penguin, 1979.
- HUME, David. An enquiry concerning the principles of morals. Trad. argentina: Investigación sobre la moral. Editorial Losada, 1945.
- IRWIN, Terence. Plato's ethics. Oxford U. P., 1995.
- JAMES, William. The principles of psychology. Trad. argentina: Editorial Glem, 1945.
- KANT, Immanuel. Théorie et pratique. Vrin, 1990.  
Fundamentação da metafísica dos costumes. Edições 70, 1986.
- KAVKA, Gregory. Hobbes's war of all against all. *Ethics*, vol. 93, n° 2, pp. 291-310.  
Right reason and natural law in Hobbes's ethics. *The monist*, vol. 66, n° 1, pp. 120-133.

- KELSEN, Hans. O problema da justiça. Martins Fontes, 1991.
- KOYRÉ, Alexandre. Du monde clos à l'univers infini. Trad. portuguesa: Gradiva, s/d.  
Études galiléennes. Trad. portuguesa: Edições Dom Quixote, 1986.
- LANDIM FILHO, Raul. Análise da noção de objeto e representação em uma filosofia da consciência e em uma filosofia lógico lingüística. Dados, vol. 33, n° 3, 1990, pp. 489-517.  
Evidência e verdade no sistema cartesiano. Edições Loyola, 1992.
- LYONS, David. Ethics and the rule of law. Trad. brasileira: As regras morais e a ética. Papirus, 1990.
- LUX, Keneth. Adam Smith's mistake. Trad. brasileira: Nobel, 1993.
- MACINTYRE, Alasdair. Whose justice? Which rationality? Trad. brasileira: Edições Loyola, 1991.
- MACKIE, John. Ethics: inventing the right and the wrong. Penguin, 1977.  
Sidgwick's pessimism. Philosophical Quarterly, 1976, pp. 317-327.
- MARCHETTI, Valmor. Risco e decisão em investimento produtivo. Editora da UFRGS, 1995.
- MONZANI, Luiz Roberto. Desejo e prazer na idade moderna. Editora da Unicamp, 1995.
- MULHOLLAND, Leslie. Egoism and morality. The Journal of Philosophy, vol. 86, n° 10, 1989, pp. 542-550.
- NAGEL, Thomas. The possibility of altruism. Princeton U. P., 1970.  
Hobbes's concept of obligation. Philosophical Review, vol. 68, 1959, pp. 68-83.  
Rawls on justice. In DANIELS (ed.) Reading Rawls.
- NINO, Carlos S. Ética y derechos humanos: un ensayo de fundamentación. Editorial Ariel, 1989.
- NOZICK, Robert. Anarchy, state and utopia. Trad. brasileira: Jorge Zahar Editor, 1991.  
Philosophical explanations. The Belknap Press of Harvard U. P., 1981.  
The nature of rationality. Princeton U. P., 1995.
- O'BRIEN, D. P. The classical economists. Trad. espanhola: Alianza Editorial, 1989.
- OLASO, Ezequiel de. Thomas Hobbes y la recta razón. Manuscrito, vol. 4, n° 1, 1980, pp. 29-35.
- PANICHAS, George E. Hobbes, prudence and basic rights. Noûs, vol. 22, 1988, pp. 555-571.
- PARFIT, Derek. Prudence, morality and the prisoner's dilemma. In ELSTER (ed.) Rational choice.
- PLATÃO. Protágoras. Editorial Inquérito, s/d.

- A República. Fundação Calouste Gulbenkian, 1987.
- Obras Completas. Aguilar, 1990. 2ª edición.
- POLYA, G. How to solve it: a new aspect of mathematical method. Princeton U. P., 1948.
- POPPER, Karl. The open universe. Trad. portuguesa: Publicações Dom Quixote, 1988.
- Realism and the aim of science. Trad. portuguesa: Dom Quixote, 1987.
- PUTNAM, Hilary. Renewing philosophy. Trad. espanhola: Cómo renovar la filosofía. Ediciones Cátedra, 1994.
- RABOSSI, Eduardo. Hobbes y la filosofía del lenguaje. Manuscrito, vol. 4, nº 1, 1980, pp. 21-28.
- Naturaleza humana y moralidad. Revista Latinoamericana de Filosofía, vol. 6, nº 2, julio 1980, pp. 119-136.
- RAPOPORT, Anatol. Fights, games and debates. Trad. brasileira: Editora da UNB, 1980.
- Uso e abuso da teoria dos jogos. In HARDIN, Garrett (org.) A ciência social num mundo em crise: textos do Scientific American. EDUSP & Editora Perspectiva, 1973.
- RAWLS, John. A theory of justice. The Belknap Press of Harvard U. P., 1971.
- RAZ, Joseph (ed.) Practical reasoning. Oxford U. P., 1978.
- RESCHER, Nicholas. Rationality and moral obligation. Synthese, vol. 72, 1987, pp. 29-43.
- RICHARDS, David. Moral rationality. Synthese, vol. 72, 1987, pp. 91-101.
- ROBINSON, Daniel. The enlightened machine: an analytical introduction to neuropsychology. Trad. brasileira: E. P. U. & EDUSP, 1977.
- ROBINSON, Joan. Economic philosophy. Penguin, 1964.
- ROBINSON, Joan & Eatwell, John. An introduction to modern economics. Trad. mexicana: Fondo de Cultura Económica, 1982.
- ROHDEN, Valério. (coord.) Racionalidade e ação. Editora da UFRGS & Goethe-Institut, 1992.
- (coord.) Ética e política. Editora da UFRGS & Goethe-Institut, 1993.
- ROUSSEAU, J.-J. Du contrat social. Larousse, 1973.
- RUNCIMAN, W. G. & SEN, A. K. Games, justice and the general will. Mind, vol. 74, 1965, pp. 554-562.
- RYLE, Gilbert. Dilemmas. Trad. brasileira: Martins Fontes, 1993.
- SALMON, Wesley. Logic. Trad. brasileira: Ed. Guanabara, 1987. 6ª edição.
- SCHELLING, Thomas C. Micromotives and macrobehavior. Trad. espanhola: Fondo de Cultura Económica, 1989.
- SCHOPENHAUER, Arthur. Über das Fundament der Moral. Trad. brasileira: Martins Fontes, 1995.

- SEARLE, John R. Intentionality: an essay in the philosophy of mind. Trad. brasileira: Intencionalidade. Martins Fontes, 1995.
- SEN, Amartya. Behaviour and the concept of preference. In ELSTER (ed.) Rational choice. Collective choice and social welfare. North-Holland, 1979.
- Rational fools: a critique of the behavioral foundations of economic theory. Philosophy & Public Affairs, vol. 6, n° 4, 1977, pp. 317-344.
- On ethics & economics. Basil Blackwell, 1987.
- SEN, Amartya & WILLIAMS, Bernard. (eds.) Utilitarianism and beyond. Cambridge U. P. & Editions de la Maison des Sciences de L'homme, 1991.
- SIDGWICK, Henry. The methods of ethics. Hackett, 1981.
- SIMON, Herbert. Reason in human affairs. Trad. portuguesa: Gradiva, 1989.
- SIMONSEN, Mário Henrique. Ensaio analítico. Editora da Fundação Getúlio Vargas, 1994.
- SINGER, Peter. Practical ethics. Cambridge U. P., 1993. 2nd. edition. Trad. brasileira: Martins Fontes, 1994.
- (ed.) A companion to ethics. Blackwell, 1991.
- SMITH, Adam. The theory of moral sentiments. Liberty Classics, 1982.
- SPINOZA, Baruch. Ética. Ediouro, s/d.
- STERBA, James. Justifying morality: the right and the wrong ways. Synthese, vol. 72, 1987, pp. 45-69.
- STRAUSS, Leo. The political philosophy of Thomas Hobbes: its basis and its genesis. The University of Chicago Press.
- TARSKI, Alfred. Introduction to logic and to the methodology of deductive sciences. 2<sup>nd</sup> edition. Trad. francesa: Introduction a la logique. Gauthier-Villars & E. Nauwelaerts, 1960.
- TAVARES, José Antônio G. Dois paradigmas de ética pública na teoria liberal. In ROHDEN (coord.) Ética e política.
- THOMAS, Laurence. Doing justice to egoism. The journal of philosophy, vol. 86, n° 10, 1989, pp. 551-552.
- TORRES, João Carlos Brum. Discutindo a lei de Gérson. In ROHDEN (coord.) Racionalidade e ação.
- TUGENDHAT, Ernst. Probleme der Ethik. Trad. castelhana: Editorial Crítica, 1988.
- URMSON. Aristotle's ethics. Blackwell, 1988.
- The goals of action. In RAZ (ed.) Practical reasoning.
- WALKER, Ralph. Kant tem uma filosofia moral? Filosofia Política 5, L & PM Editores, 1989.

- WALLACE, R. Jay. How to argue about practical reason. Mind, vol. 99, July 1990, pp. 355-385.
- WARNOCK, G. J. The object of morality. Methuen, 1971.
- WARNOCK, Mary. The uses of philosophy. Trad. brasileira: Papirus, 1994.
- WILLIAMS, Bernard. Ethics and the limits of philosophy. Harvard U. P., 1985.
- Problems of the self. Cambridge U. P., 1973.
- Moral luck: philosophical papers 1973-1980. Cambridge U. P., 1981.
- Morality: an introduction to ethics. Cambridge U. P. The Canto edition, 1993.
- WITTGENSTEIN, Ludwig. Tractatus Logico-Philosophicus. EDUSP, 1994. 2ª edição.
- Wittgenstein's lectures and conversations. Trad. portuguesa: Cotovia, 1991.
- WOLFF, Robert P. The poverty of liberalism. Trad. brasileira: Ed. Paz e Terra, 1990.