BRUNO LOPES VIEIRA

VALORES, MEDIDAS E ROBÔS:
A INCOMENSURABILIDADE DE VALORES NAS DECISÕES TOMADAS POR INTELIGÊNCIAS ARTIFICIAIS

Londrina
2022

BRUNO LOPES VIEIRA

VALORES, MEDIDAS E ROBÔS:
A INCOMENSURABILIDADE DE VALORES NAS DECISÕES TOMADAS POR INTELIGÊNCIAS ARTIFICIAIS

Dissertação apresentada ao Programa de Graduação *Stricto Sensu* em Filosofia da Universidade Estadual de Londrina – UEL, como requisito final à obtenção do título de Mestre em Filosofia.

Orientadora: Profª. Drª. Andrea Luisa Bucchile Faggion

Londrina
2022

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

V658v Vieira, Bruno Lopes.
Valores, medidas e robôs : a incomensurabilidade de valores nas decisões tomadas por inteligências artificiais / Bruno Lopes Vieira. - Londrina, 2022.
76 f. : il.

Orientador: Andrea Luisa Bucchile Faggion.
Dissertação (Mestrado em Filosofia) - Universidade Estadual de Londrina, Centro de Letras e Ciências Humanas, Programa de Pós-Graduação em Filosofia, 2022.
Inclui bibliografia.

1. Incomensurabilidade de valores - Tese. 2. Comparações - Tese. 3. Inteligência artificial - Tese. 4. Teoria da decisão - Tese. I. Faggion, Andrea Luisa Bucchile. II. Universidade Estadual de Londrina. Centro de Letras e Ciências Humanas. Programa de Pós-Graduação em Filosofia. III. Título.

CDU 1

BRUNO LOPES VIEIRA

VALORES, MEDIDAS E ROBÔS:
A INCOMENSURABILIDADE DE VALORES NAS DECISÕES TOMADAS POR INTELIGÊNCIAS ARTIFICIAIS

Dissertação apresentada ao Programa de Graduação *Stricto Sensu* em Filosofia da Universidade Estadual de Londrina – UEL, como requisito final à obtenção do título de Mestre em Filosofia

BANCA EXAMINADORA

Orientadora: Profª. Drª. Andrea Luisa Bucchile Faggion
Universidade Estadual de Londrina – UEL

Prof. Dr. Charles Feldhaus
Universidade Estadual de Londrina – UEL

Prof. Dr. Thomas da Rosa de Bustamante
Universidade Federal de Minas Gerais – UFMG

Londrina, 07 de julho de 2022.

CENTRO DE LETRAS E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

ATA DE DEFESA DE DISSERTAÇÃO

As sete dias do mêsde julho do ano de dois mil e vinte e dois, às nove horas, em Sala Virtual, do Centro de Letras e Ciências Humanas, desta Universidade, reuniu-se a Banca Examinadora homologada pelo Programa de Pós-Graduação em Filosofia, composta por Dra. Andrea Luisa Bucchile Faggion, Dr. Thomas da Rosa de Bustamante e Dr. Charles Feldhaus. A reunião teve por objetivo julgar o trabalho do estudante Bruno Lopes Vieira, sob o título "VALORES, MEDIDAS E ROBÔS: A INCOMENSURABILIDADE DE VALORES NAS DECISÕES TOMADAS POR INTELIGÊNCIAS ARTIFICIAIS". Os trabalhos foram abertos pela professora Andrea Luisa Bucchile Faggion. A seguir, foi dada a palavra ao estudante, que apresentou seu trabalho remotamente. Cada examinador arguiu o Mestrando, com tempos iguais de arguição e resposta. Terminadas as arguições, procedeu-se o julgamento do trabalho, sendo que todos os professores enviaram simultaneamente os formulários de avaliação, os quais foram impressos e anexados à presente ata. A Banca Examinadora concluiu pela **aprovação** do trabalho. Nada mais havendo a tratar, foi lavrada a presente ata, que vai assinada pela Presidente da Banca Examinadora.

A estudante deverá reformular seu trabalho no prazo de _____ dias: ( ) SIM (X) NÃO

Se houver alteração no título do trabalho, informar o novo título abaixo:

_____

_____

_____

OBS.: Este documento não deve conter rasuras ou corretivo e deve ser preenchido de forma legível

Londrina, 07 de julho de 2022.

PRESIDENTE

Dra. Andrea Luisa Bucchile Faggion UEL

TITULARES

Dr. Thomas da Rosa de Bustamante UFMG

Dr. Charles Feldhaus UEL

VIEIRA, Bruno Lopes. **Valores, medidas e robôs:** a incomensurabilidade de valores nas decisões tomadas por inteligências artificiais. 2022. 76 f. Dissertação (Mestrado em Filosofia) – Universidade Estadual de Londrina, Londrina. 2022.

RESUMO

A área denominada teoria da decisão sofre influências de diversos campos do conhecimento como filosofia, economia, psicologia, computação, entre outras. Em suma, ela busca compreender como decisões *deveriam* ser tomadas, assim como elas *realmente* são tomadas. Uma das questões investigadas é se o fenômeno da incomensurabilidade de valores afeta a possibilidade da decisão racional, ou seja, se é possível fundamentar uma escolha quando as opções não possuem uma medida comum entre elas, para que possam ser comparadas. Neste trabalho, foi analisada a estrutura das comparações e o seu papel na tomada de decisões racionais para que, adiante, pudesse ser verificado se a ausência de medida comum afetaria a comparabilidade das opções. Adiante, foram vistas teorias que lidam com as denominadas escolhas difíceis, a fim de se buscar uma saída para o imbróglio da incomensurabilidade. Por fim, considerando os resultados dessas teorias, foi verificado se estes poderiam ser estendidos a agentes não-humanos, no caso, inteligências artificiais. Concluiu-se que, devido as características necessárias para operar escolhas difíceis, a incomensurabilidade de valores representa um desafio não superado para a inteligência artificial.

Palavras-chave: comparações; incomensurabilidade de valores; inteligência artificial; paridade; teoria da decisão.

VIEIRA, Bruno Lopes. **Values, measures, and robots: the value incommensurability in artificial intelligences decision making**. 2022. 76 p. Dissertation (Master's in Philosophy) – Londrina State University, Londrina, 2022.

ABSTRACT

The decision theory is influenced by several other fields of knowledge such as philosophy, economics, psychology, computing, among others. Roughly, it seeks to understand *how* decisions should be made, as well as how they are *actually made*. One of the pursued questions is whether the phenomenon of value incommensurability affects the possibility of rational decision, that is, if it is possible to justify a choice when the options don't have a common measure among them, so that they can be compared. In this thesis, the structure of comparisons and their role in rational decision-making were analyzed so that, further on, it could be verified whether the absence of a common measure would affect the comparability of options. In sequence, theories were seen that deal with the so-called hard choices, in order to find a way out of the incommensurability imbroglio. Finally, considering the results of these theories, it was verified whether they could be extended to non-human agents, in this case, artificial intelligences. It was concluded that, due to the characteristics necessary to operate hard choices, the incommensurability of values represents an unsurpassed challenge for artificial intelligence.

Palavras-chave: artificial intelligence; comparisons; decision theory; value incommensurability; parity.

SUMÁRIO

INTRODUÇÃO

As escolhas têm um papel crucial na vida humana. Todos os dias, tomamos dezenas ou centenas de decisões. Até a mera inércia pode ser entendida como a decisão de não se movimentar. São inúmeros os exemplos de como a humanidade se interessa pelo ato de escolher.

No universo literário, há um vasto material sobre as dificuldades inerentes às escolhas, assim como suas inexoráveis consequências. William Shakespeare, em *Hamlet* (1603), ilustra de forma brilhante a complexidade, muitas vezes negligenciada, de uma escolha aparentemente simples. Em *O Senhor dos Anéis* (1954), a escolha de Isildur em manter o Anel de Sauron, em vez de destruí-lo na Montanha da Perdição, condenou a si e a milhares de indivíduos ao tormento. No romance *A Escolha de Sofia* (1979), as alternativas oferecidas à protagonista demonstram que algumas escolhas têm um destino inevitavelmente trágico, independente da opção escolhida.

No cinema, em *Indiana Jones e a Última Cruzada* (1989), o herói precisa escolher entre o cálice verdadeiro e o cálice falso — escolha que definirá seu destino. Em *The Matrix* (1989), a clássica cena em que Neo precisa escolher entre a pílula vermelha, que o faria despertar para o mundo real, e a pílula azul, que apagaria suas memórias e o levaria para uma vida artificial confortável, é amplamente citada na cultura popular como um dilema entre escolher a dor da verdade e o conforto da ignorância. Ainda hoje, o termo *red pill* é usado para significar uma grande epifania.

No campo filosófico, as questões inerentes às escolhas também tiveram grande atenção. Aristóteles, em *Ética à Nicômaco* (aproximadamente 300 a.C.), talvez tenha sido o primeiro a dedicar tanto esforço ao tema. Pelo que chamou de *phronêsis*, ou sabedoria prática, o filósofo grego designou a virtude do pensamento prático, traçando diretrizes para o exercício prudente da tomada de decisão. Mais adiante, mesmo aqueles que não trataram especificamente sobre o tema, reconheceram a sua importância. Jean-Paul Sartre (1966, p. 21), ao cunhar a célebre frase, central ao conceito do existencialismo, *“l'existence*

précède l'essence"[1], tinha em mente o papel transformador das decisões e como elas criam valores e determinam o sentido da vida dos indivíduos.

Outras áreas do conhecimento como a teologia, a matemática, a ciência econômica, a administração, o *marketing*, a medicina, o direito e a psicologia também tratam — cada uma em seu recorte teórico — do papel essencial que as decisões desempenham em seus objetos de estudo. Nos dias atuais, o fato de que os livros de autoajuda costumam figurar entre os gêneros literários mais vendidos no mundo, assim como o fenômeno crescente do *life coaching*, podem indicar que, cada dia mais, as pessoas buscam respostas para a pergunta: quais decisões devo tomar?

Na investigação filosófica, um certo tipo de situação de escolha é particularmente problemática: aquela em que os elementos a serem comparados não compartilham de uma mesma medida comum em que se possa comensurá-los. Tal fenômeno chamaremos de incomensurabilidade. A possibilidade da decisão justificada tem, portanto, sido debatida em situações desse tipo, e muito tem sido escrito em busca da resposta para o problema de decidir diante de incomensuráveis.

Contudo, indiferente ao fato de que a resposta sobre as melhores decisões ainda não está bem estabelecida para a humanidade, a ciência da computação e a tecnologia da informação avançaram de tal maneira que, atualmente, sistemas binários tomam milhares de decisões de forma automatizada, pelos mecanismos de *machine learning* e inteligência artificial. Como qualquer espécie de decisão, as decisões orientadas pela inteligência artificial têm o potencial de influenciar drasticamente a vida humana, porém, com uma diferença substancial: tais decisões independem, em certa medida, da ação humana.

Embora o estudo sobre decisões baseadas em inteligências artificiais seja um campo de estudo recente, é algo que, assim como as decisões humanas, foi objeto de reflexão no mundo artístico. O cenário apocalíptico causado por uma rebelião das máquinas foi enredo de diversas obras culturais. Issac Asimov foi, provavelmente, um dos mais notáveis escritores sobre a possibilidade da

[1] "A existência precede a essência".

dominação do homem pelas máquinas. Seus livros como *Eu, Robô* e *Fundação*, apesar de classificados como ficção científica, possuem um nível de detalhamento tão profundo que poderia se confundir com um tratado teórico sobre ética e tecnologia. No mesmo sentido, outros ícones da cultura popular como *Frankenstein*, *Odisseia no Espaço*, *O Exterminador do Futuro* e *Battlestar Galactica* dão suporte para a plausibilidade da aflição humana em relação às máquinas.

Mesmo que isso soe distante ou até conspiracionista a alguns, o fato é que fenômenos como a autorreplicação [2], antes reservados às ciências biológicas e a modelos matemáticos, hoje é estudado pela nanotecnologia (LI et. al, 2010) com possíveis implicações na robótica. Além disso, figuras públicas da tecnologia e da ciência, como Stephen Hawkins, Elon Musk[3] e Bill Gates[4] já se pronunciaram publicamente alertando as pessoas sobre as possíveis ameaças da inteligência artificial ao futuro da humanidade, enfatizando a importância da condução de pesquisas sérias sobre o comportamento de sistemas automatizados. Nesse sentido ainda, a Universidade de Oxford inaugurou recentemente (2021), o *Institute for Ethics in AI*, dedicado ao estudo da ética em inteligência artificial.

Tendo em vista que o escopo das decisões tomadas por algoritmos não abrangem somente valores comensuráveis, o objetivo deste trabalho é compreender o impacto que essa questão complexa da teoria da decisão, qual seja, a incomensurabilidade de valores, pode ter nas decisões automatizadas tomadas por algoritmos inteligentes, e se é possível que os parâmetros utilizados para que se considerar racionalmente justificável as decisões humanas, possa ser utilizado para guiar as decisões tomadas por máquinas.

Iniciarei analisando as principais teorias de justificação racional, a fim de responder, do ponto de vista prático, o que é uma decisão racionalmente justificada. Para isso, visitarei os argumentos oferecidos pelos filósofos que

---

[2] A capacidade de um sistema dinâmico reproduzir uma cópia idêntica de si. Na biologia, podemos citar a divisão celular do DNA na reprodução.

[3] *Vide:* FUTURE OF LIFE INSTITUTE. An open letter. **Research priorities for robust and beneficial artificial intelligence**. Disponível em: *https*://futureoflife.org/ai-open-letter/. Acesso em: 17 jul. 2022.

[4] *Vide:* BBC. Microsoft's Bill Gates insists AI is a threat. **BBC News**, 29 de janeiro de 2015. Disponível em: https://www.bbc.com/news/31047780. Acesso em: 17 jul. 2022.

buscaram teorizar sobre isso. Nesse intento, no primeiro capítulo buscarei compreender se a comparabilidade entre os elementos é uma condição necessária para a justificação racional. Isso porque poderia se pensar que nem sempre precisamos comparar os elementos para tomar uma decisão justificada. Por exemplo, diante de uma situação em que uma determinada escolha é a melhor e a outra é a exigida pela lei, não seria justificável optar por esta, ainda que a comparação apontasse que aquela é a melhor? Parece, então, que a justificação não dependeria necessariamente da comparação, pois bastaria recorrer, em alguns casos, a uma regra ou diretriz de *status* superior que não se utilizam da comparação entre as alternativas para chancelar o que é racionalmente justificável.

De qualquer forma, ainda que a racionalidade não exija que a comparabilidade esteja presente em qualquer decisão justificada, é razoável imaginar que a comparabilidade desempenhe, em algumas situações, um papel fundamental na resposta justificada. Por conta disso, no segundo capítulo irei assumir a importância da comparabilidade e me aprofundar no impacto causado pelo fenômeno da incomensurabilidade — entendida aqui como a ausência de medida comum entre as opções em jogo — na possibilidade de se comparar as opções. Em outras palavras, irei buscar uma resposta à pergunta: se dois elementos não possuem uma medida comum entre eles, é possível que possamos compará-los de forma racional? Será estudada, especialmente, a resposta dada pela teoria da paridade, de Ruth Chang, que aponta uma quarta relação comparativa de valor, em que se encontram as opções incomensuráveis, e permite a operabilidade da escolha racional por meio de elementos subjetivos do agente.

No último capítulo, uma premissa inicial será assumida: a de que o principal propósito da tecnologia de inteligência artificial é a de buscar, ou ao menos maximizar, a ação racional. Diante disso, será analisada a dinâmica das preferências humanas que, como sustentado, podem operar a escolha entre bens incomensuráveis. Ao tratar das preferências, será dada ênfase àquelas mais valiosas à humanidade, chamadas genericamente de valores. Assim, será abordada a possibilidade de os algoritmos conhecerem e definirem os valores humanos para que possam tomar decisões com base neles. Por fim, irá se

concluir não há um prognóstico confiável acerca da possibilidade de a inteligência artificial conhecer os valores humanos, especialmente por conta da individualidade e do dinamismo, intrínsecos à natureza das preferências.

1 A JUSTIFICAÇÃO DA ESCOLHA

All we have to decide is what to do with the time that is given us.

Gandalf

A vida humana é composta por uma miríade de possíveis ações que podemos executar. Como diz o ditado popular: a vida é feita de escolhas. Contudo, ao menos intuitivamente, nos referimos com certa frequência à justificação dessas escolhas, no sentido de haver escolhas justificáveis e injustificáveis.

A possibilidade da escolha justificada entre duas ou mais alternativas imprescinde, segundo parte dos teóricos[5], da comparabilidade entre as opções, ou seja, só é possível que uma opção seja escolhida em detrimento de outras de forma justificada se for possível comparar, de alguma forma, o conteúdo das alternativas. Um alegado problema para a comparação entre alternativas é que a ausência de medida comum — a incomensurabilidade — entre os valores carregados pelas opções acarretaria a impossibilidade da comparação, o que, por sua vez, implicaria na impossibilidade de uma escolha justificada.

Talvez por conta de o tema estar em debate aprofundado há um tempo relativamente curto do ponto de vista investigativo-filosófico, existem muitas ideias ligadas à incomensurabilidade na literatura. Muitas dessas ideias são incompatíveis entre si e, conforme irá se argumentar, equivocadas em relação ao que deveria se entender sobre o que é, de fato, ser um valor incomensurável. Por conta disso, este capítulo irá definir um chão comum em relação aos termos utilizados nesse trabalho, explorando algumas ideias relevantes ao objeto da pesquisa, assim como entender o que se quer dizer quando se diz que uma opção é justificável.

[5] Filósofos utilitaristas, como Jeremy Bentham (2000) e John Stuart Mill (2009), por exemplo, consideram que a comparabilidade seja essencial, pois ela está no núcleo dessa teoria. Uma vez que valores distintos podem ser resumidos a instâncias de algum outro valor como prazer ou felicidade, então a justificação da escolha se encontra na opção que apresenta maior quantidade desse valor supremo. Outros filósofos que não se filiam a correntes utilitaristas também defendem a importância da comparabilidade na justificação como Donald Regan (1997) e Ruth Chang (1997, 1998, 2015).

1.1 JUSTIFICAÇÃO EPISTEMOLÓGICA

Quando falamos em justificação, muitas ideias podem aparecer. Por exemplo, dentro da investigação epistemológica, a justificação guarda relação com as crenças que carregamos. Justificar, no sentido epistêmico do termo, é estabelecer uma relação de legitimidade, ou de verdadeiro, com a crença que dá causa a uma ação.

Tendo em vista que é contraditório que tomemos alguma ação com base em algo que não acreditamos[6], todas as nossas ações estão, de certa forma, conectadas a alguma crença ou conjunto de crenças. Se alguém comemora seus aniversários na mesma data, pois pensa que nasceu nessa data, uma vez que existe uma certidão de nascimento e seus pais também afirmam a mesma data, isso implica em algumas crenças. Ao menos, a pessoa acredita que aquele documento é verídico e que não houve erro ou manipulação da informação escrita. Além disso, também acredita que seus pais não estão mentindo e nem equivocados sobre a sua data de nascimento.

Um típico problema enfrentado pelos teóricos da justificação epistemológica é a questão da inferência doxástica[7]. Esse problema assume que o nosso sistema de justificação é uma longa cadeia de crenças em que cada crença se justifica por uma crença anterior, regredindo assim sucessivamente. Utilizando o exemplo dado, se a pessoa quisesse verificar a autenticidade da certidão, ela poderia ir até o cartório de documentos e checar com o tabelião se aquela certidão foi emitida daquela forma. Caso confirmasse que sim, ainda precisaria acreditar que não houve algum erro de digitação do cartorário que emitiu o documento à época do seu nascimento. Inevitavelmente, para os

[6] É possível agir com base em um conteúdo que se reputa ilegítimo. Por exemplo, um aluno pode estudar para uma prova sem acreditar que aquele conteúdo tem alguma relevância na sua vida e que, portanto, seria melhor que ele estivesse envolvido com outra atividade. Contudo, se ele decide estudar e estuda, alguma crença está envolvida nessa ação. Pode ser que ele acredite que aquele conteúdo irá cair na prova e que aquela prova tem relevância para que ele seja aprovado na disciplina. Se na situação concreta, houvesse apenas crenças contrárias ao ato de estudar, seria contraditório que o agente agisse nesse sentido. Portanto, toda a ação está, em alguma medida, direta ou indiretamente, conectada a alguma crença.

[7] Existem correntes contrárias, como o fundacionalismo, que nega que as crenças sejam sempre justificadas por outras crenças infinitamente, de modo que exista uma crença fundamental (ou fundacional) que é sólida o suficiente para justificar as crenças a partir dela. Uma versão do fundacionalismo — o fundacionalismo cartesiano — foi amplamente defendida por René Descartes (2008).

proponentes da inferência doxástica (POLLOCK, 1986, p. 19) essa pessoa chegaria em um ponto da cadeia de crenças em que se veria diante do problema cético da regressão infinita.

Em síntese, o problema seria o fato de que como a regressão não tem fim, não há uma crença inicial que sirva de sustento para as crenças seguintes da cadeia. Logo, se não é possível justificar por meio de uma crença inicial, não seria possível justificar qualquer crença seguinte, de modo que o conhecimento seria impossível. Entretanto, esse não é um problema que irei enfrentar aqui. O sentido de justificação que irei me referir, embora possa ter alguma sobreposição com os problemas epistemológicos, não o terão como objeto principal.

1.2 JUSTIFICAÇÃO MORAL

Outra forma de entendermos a justificativa é pelo ponto de vista estritamente ético. Nesse sentido, estamos buscando uma resposta para a legitimidade moral das nossas ações. Pergunta-se, nesses casos, coisas como “devemos ajudar uma pessoa em estado de necessidade quando nós mesmos estamos em estado de necessidade?”. O que se quer é buscar uma resposta que justifique moralmente um ou outro curso de ação. Nesse caso, a justificativa buscaria legitimar ou a ação que ajuda o necessitado ou a ação que não ajuda o necessitado (ou ambas).

Para fundamentar a natureza das razões que justificam a ação, David Copp (1995) sustenta que o que determina a racionalidade das ações são as necessidades objetivas e valores subjetivos do agente. Jonathan Dancy (2004) tem uma visão parecida, porém, ao invés de distinguir desejos de necessidades, distingue entre razões peremptórias e razões atraentes. Peremptórias são aquelas que geram uma alegação de dever. Atraentes fazem uma ação ser racionalmente atraente, mas sem ser o caso de se tornar um dever fazê-la, mesmo sendo as únicas razões no caso.

Nesse sentido, Joshua Gert (2007) apresenta uma teoria da justificação moral que separa as forças das razões em duas: a força requisitante e a força justificadora. Enquanto as razões requisitantes exigem a execução de uma ação, as razões justificadoras apenas justificam — em sentido estrito — uma

determinada ação. Assim, diante de uma situação de escolha, é possível que o sujeito esteja justificado tanto em uma opção quanto em outra, de modo que não exista apenas uma única opção justificada.

A teoria de Gert pode ser entendida em relação ao sacrifício do agente. Por exemplo, imagine uma mãe solteira que trabalha em tempo integral para sustentar os seus filhos. Como a maior parte de seu salário é gasto com a subsistência mínima dela e dos filhos, ela pensa em usar seu único tempo livre à noite para fazer um curso profissionalizante que pode lhe proporcionar uma melhor renda no futuro. A ação de fazer o curso, que sacrifica o tempo com os filhos, tem força justificante, pois é racional investir tempo em prol do bem-estar futuro da família. Porém, a mesma ação não tem força requisitante, pois o fato de ela não fazer o curso não seria o suficiente para censurá-la como irracional.

Razões morais podem perfeitamente ser parte do balanço de razões de um indivíduo — e frequentemente são. Dessa forma, o estudo da justificação moral pode ter uma sobreposição com os conceitos de justificação prática que irei abordar adiante. Contudo, fiz questão de criar uma cisão, para que a essência daquilo que foi observado nas teorias acima não fosse confundida conceitualmente com o propósito deste trabalho. Assim, para o bem do refinamento do objeto, podemos entender que as razões morais podem fazer parte da justificação prática, desde que elas sejam entendidas como auxiliares[8] para o propósito do agente.

1.3 JUSTIFICAÇÃO PRÁTICA E COMPARAÇÕES

Este trabalho se ocupará unicamente da justificativa prática, em outras palavras, se preocupará com a legitimidade[9] dos fundamentos que justificam um determinado curso de ação, buscando um sentido prático à pergunta: “o que justifica uma escolha?”. A busca aqui terá, portanto, um sentido objetivo. Trata-se, portanto, do *fundamento* de uma escolha, como uma garantia que é

[8] Não me comprometo aqui como o conceito de razões auxiliares de Joseph Raz (2002, p. 34-35), mas ele pode ajudar a elucidar o meu ponto.
[9] Novamente, não no sentido estritamente moral e nem no sentido epistemológico.

endereçada ao sujeito para que este opte por determinada alternativa, independente do seu estado epistêmico (CHANG, 1998, p. 1576).

No estudo da razão prática, é tipicamente entendido que a justificativa de uma escolha é influenciada pela força, ou peso, de suas razões. Assim, uma ação será justificada se for baseada em uma ou mais razões fortes, desde que não haja uma ou mais razões de maior força para deixar de agir. Diante de opções que ofereçam razões para diferentes cursos de ação, a escolha justificada seria aquela cujas razões para agir fossem mais fortes do que a opção concorrente, vencendo assim o chamado balanço de razões.

Se a justificação da decisão depende de como avaliamos (ou pesamos) a força das razões conflitantes que se aplicam ao caso, então parece depender de algum tipo de comparação entre essas razões. Ruth Chang (1998, p. 1569) afirma que há uma sabedoria popular que assume a impossibilidade de justificar a escolha entre opções que não podem ser comparadas. Disso, implica-se logicamente que a comparação é uma condição necessária para a justificação da escolha (CHANG, 1998, p. 1569).

Embora realmente exista um certo apelo intuitivo a respeito do papel justificador das comparações, esse papel vem sendo questionado. Em outras palavras, o valor teórico do que chamaremos de *comparativismo*[10] está em debate. Em primeiro lugar, ainda que aceitemos que as comparações tenham um papel justificador, isso não implica que a comparabilidade seja uma condição necessária para a justificação racional. Ou seja, pode ser que a razão chancele um elemento comparativo como fonte de justificação, mas também que legitime elementos não-comparativos para os mesmos fins.

Nesse sentido, existem alguns pensadores que sustentam que a comparabilidade não é necessária para justificar a escolha. Entre eles, há aqueles que pensam que, quando houver uma falha na comparabilidade, um critério não-comparativo é legítimo para justificar a escolha e outros que entendem que critérios não-comparativos são legítimos mesmo quando a possibilidade de comparação esteja presente.

[10] Neste trabalho, o termo *comparativismo* será entendido amplamente como a visão que considera que a comparação é uma condição necessária para que haja a justificação da escolha.

Joseph Raz (1997, p. 110) defende que, caso estejamos diante de opções incomparáveis, a vontade do agente pode justificar a escolha racional. James Griffin (1997, p. 35) defende que a prudência e o consenso moral e legal podem ajudar a formar as normas que nos fornecem padrões para justificar a escolha. Para ambos, a comparação tem um papel privilegiado, mas não exclusivo, na justificação da escolha.

Para outros autores (ANDERSON, 1997, p. 90; GRIFFIN, 1997, p. 35, 50; MILLGRAM, 1997, p. 53-66; STOCKER, 1997, p. 196; TAYLOR, 1997, p. 170), a comparabilidade não desempenha um papel privilegiado ou sequer pode ter um papel justificador. Para estes, os fatos concretos sobre uma opção ou a forma como os bens se articulam em determinados contextos, são critérios bastantes para a justificação racional da escolha.

1.4 COMPARABILIDADE

Antes de avançarmos a discussão sobre a questão de se a justificação racional depende necessariamente de alguma comparação, é importante estabelecer os critérios relevantes para o conceito de comparabilidade.

A comparação é um método de verificação com o intuito de se extrair um resultado de como dois ou mais itens se relacionam entre si a respeito de uma determinada característica. Nesse sentido, Chang (2013b, p. 8) aponta duas características centrais da comparabilidade, assim, uma relação comparativa sempre irá possuir "uma relação *positiva* de valor de acordo com a qual eles possam ser classificados" e procederá "com respeito a um 'elemento relativo[11]' avaliativo".

[11] Inicialmente, Chang chamou esse atributo, o qual traduzi como "elemento relativo", de *"covering value"* (1997). Em publicações mais recentes, passou a adotar o termo *"covering consideration"* (2002b, 2013) de forma mais ampla e o termo *"choice value"* no contexto de escolhas (2008). Escolhi o termo "elemento" intencionalmente para não me restringir às considerações que envolvam necessariamente valores. Isso porque, neste ensaio, eu não quis me comprometer com a ampla carga semântica e as diversas teorias sobre o que se constitui um valor.

1.4.1 Elemento relativo

O elemento relativo diz respeito à consideração sob a qual se quer comparar. É aquilo que importa — o que é relevante — em uma situação comparativa particular. Assim, podemos comparar os mesmos itens, ou valores, em relação a elementos (considerações, valores, emoções etc.) diferentes.

Se compararmos, por exemplo, maçãs com laranjas, podemos proceder em relação a diversos elementos relativos. Esses elementos podem ser mais objetivos como preço, densidade nutricional, peso, quantidade delas em uma cesta, ou mais subjetivos como sabor, relevância cultural em uma comunidade, harmonização com um certo tipo de vinho, valor afetivo, entre outros.

Desse modo, as avaliações comparativas poderão ter resultados diferentes de acordo com o elemento relativo. Seguindo o exemplo, um sujeito pode escolher uma maçã ao invés de uma laranja se o elemento relativo importante na decisão for o sabor (partindo da premissa que ele gosta mais do sabor de maçãs do que de laranjas). Contudo, se o que importar nesta decisão específica for a quantidade de vitamina C (talvez o sujeito esteja gripado e acredite que uma maior ingestão de ácido ascórbico ajudará em sua recuperação), ele estará em maior conformidade com suas razões se escolher a laranja.

1.4.2 Relação de valor positiva

Uma relação de valor positiva pode ser definida como aquela cujo julgamento nos informa como dois itens se relacionam (positivamente) entre si, ao invés de como não se relacionam (HSIEH, 2021). Por exemplo, se dissermos que ir ao cinema é mais prazeroso do que ler um livro, há uma relação positiva de valor, pois não há mais nada a ser dito em relação a como às atividades se comparam, contudo, se dissermos que ir ao cinema *não* é mais prazeroso do que ler um livro, não temos um resultado definido, pois pode ser o caso de que ir ao cinema é mais prazeroso do que a leitura ou que é igualmente prazeroso. Nesse exemplo, o elemento relativo da comparação é o prazer proporcionado ao sujeito.

Quando estamos tratando de valores, é natural assumir que alguns deles possuem dimensões quantitativas e qualitativas. Por exemplo, valores abstratos como o prazer podem ser avaliados, ainda que de forma imprecisa, em aspectos quantitativos. A secreção de adrenalina na corrente sanguínea, proporcionada por um passeio em uma montanha-russa veloz, pode ser avaliada como uma grande *quantidade* de prazer para alguns indivíduos. Esses mesmos indivíduos podem considerar prazerosa uma atividade monótona como a observação pacífica de cadeias montanhosas. Ainda que menos intensa (menor quantidade de prazer), esta última não é necessariamente considerada pior do que a primeira, pois possui um aspecto qualitativo distinto — um tipo, espécie ou qualidade diferente de prazer.

Mesmo que a comparação aconteça entre itens que possam ser comparados de forma quantitativa e precisa, é possível "entender" a comparação em termos qualitativos. Imagine que alguém esteja tentando derrubar uma manga de uma mangueira, mas não a alcance com as mãos. A manga está há aproximadamente dois metros de distância da sua cabeça, então ele apanha dois galhos de árvore que encontrou no chão para que sirvam de ferramenta para seu objetivo (derrubar a manga). O primeiro galho tem um metro e se provou ineficaz para o propósito. O segundo galho, com um metro e noventa centímetros, conseguiu derrubar, após algum esforço, a manga da mangueira.

No exemplo, o aumento da dimensão quantitativa do bem implicou em um aumento proporcional da dimensão qualitativa dele, pois aumentou seu valor do tipo "alcançar a manga". Porém, se houvesse um terceiro galho, com cinco metros, o ganho em quantidade teria piorado a dimensão qualitativa do bem, pois seria ineficaz para o seu propósito. Assim, ainda que alguns bens costumem ser entendidos pelo paradigma de "quanto mais, melhor" (tipicamente, o dinheiro), isso está longe de ser verdadeiro para todos os tipos de bens. Feita a distinção entre aspectos quantitativos e qualitativos, podemos focar neste último.

Além disso, valores podem ser positivos e negativos. Usando valores abstratos como exemplos, podemos entender o amor e a amizade como positivos, e a dor e a angústia como negativos. Deixando de lado a sobreposição que essas afirmações possam ter no campo estritamente moral, e para facilitar a compreensão, trataremos aqui apenas de valores positivos, que podem ser

amplamente entendidos como aqueles que são bons para as pessoas em relação a algum aspecto.

Entre dois valores, a depender de suas cargas axiológicas, haverá um julgamento sobre a relação de valor positiva que é extraída da comparação entre eles. Assim, se um determinado valor x é bom em relação a v (em que v é o elemento relativo), por possuir uma certa "carga positiva", for comparado com um valor y, que também é bom, mas que possui uma "maior[12] carga positiva", y será considerado melhor que x, pois "melhor" é o termo comparativo tradicionalmente usado para identificar que alguma coisa é "mais boa[13]" do que outra. De igual modo, na comparação inversa, x é pior que y em relação a v. Se a carga positiva de x, em relação a v, pudesse ser aumentada de modo que se equiparasse à carga positiva de y, é razoável pensar que x seria julgado como igualmente bom a y em relação a v.

Assim, as relações de valor positivas entre as dimensões qualitativas de um valor são tradicionalmente exauridas pelas opções: melhor que, pior que, e igualmente bom. Essa limitação da extensão conceitual das relações de valor positivas foi chamada de tese da tricotomia (CHANG, 1997, p. 4). Por muito tempo, essa tese foi dada como garantida na literatura sobre o tema, como uma verdade autoevidente e não questionada. Contudo, a investigação filosófica mais recente levanta questões acerca desse limite, sugerindo que há espaço para uma quarta relação de valor. Essa discussão será abordada no capítulo seguinte, no tópico 2.4.

---

[12] Por limitações linguísticas, não pude fazer uma referência perfeita a uma ênfase no aspecto qualitativo de um bem sem recorrer a um termo quantitativo, como "maior". Isso pode causar alguma confusão, especialmente porque maior é o superlativo de grande, que possui significado quantitativo. Desse modo, maior está para "mais grande", assim como melhor está para "mais bom". No inglês, é comum o uso do *great* para tais situações, uma vez que o termo em questão pode ser usado para enfatizar aspectos quantitativos como em *"a great crowd had gathered"*, e aspectos qualitativos como em *"great times are coming"* ou *"you have a great collection of books"*. Nesse último exemplo, pode ser uma coleção magnífica de livros, sem ser quantitativamente grande. Em alguns casos, no português, é possível que se use o grande com ênfase qualitativa, no sentido de grandiosidade, como em "seu pai foi um grande homem". Contudo, no tocante à comparação entre valores, não consegui encontrar um termo mais adequado para me referir aos aspectos qualitativos sem algum prejuízo semântico.

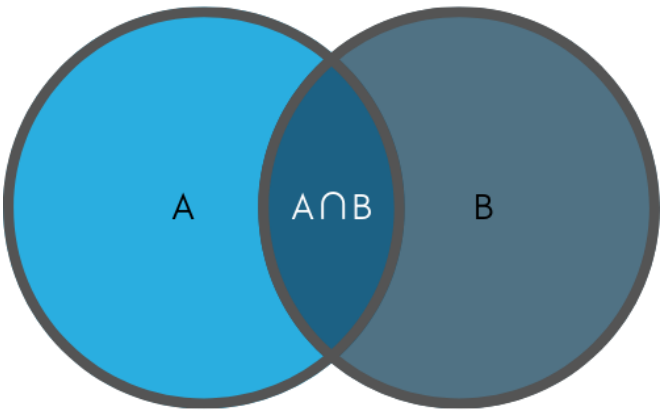
[13] O erro formal é proposital para demonstrar o que se quer dizer quando se diz que algo é melhor. Assim como na nota anterior, aqui incorro no problema de usar um termo que remete inevitavelmente a uma ideia quantitativa, o "mais". Em outras palavras, se algo é bom é porque possui uma dimensão de valor positivo. Se outra coisa que se quer comparar tiver uma dimensão de valor superior ou enfatizada, será considerada melhor.

1.4.3 Comparações *sic et simpliciter* e não-comparabilidade

A delimitação conceitual dos requisitos da comparabilidade, acima mencionados, é importante para que evitemos a típica confusão de dizer que uma coisa é *simplesmente* melhor do que outra. O item a pode ser melhor que o item b, em relação ao elemento v^1, mas pior que b, em relação ao elemento v^2. Por exemplo, um automóvel do modelo Honda Civic pode ser melhor que um do modelo Ford Fiesta em relação à tecnologia mecânica, mas pior em relação ao consumo de combustível. Quando dizemos que algo é *simplesmente* melhor, estamos pressupondo, conscientemente ou não, um elemento relativo ou um conjunto de elementos relevantes para a decisão.

De forma análoga, podemos compreender as relações comparativas pelo prisma da teoria dos conjuntos da matemática. Imagine dois conjuntos que representem dois portadores de valor quaisquer. Os elementos que esses conjuntos carregam são seus predicados. No contexto da escolha, imagine que um indivíduo está seriamente em dúvida sobre se deve se casar ou comprar uma bicicleta. O casamento e a bicicleta são portadores de valor que compartilham alguns predicados, mas não todos. Os predicados em comum entre A e B formam uma interseção entre esses conjuntos e somente nesse subconjunto A∩B é que é possível que haja comparabilidade ou incomparabilidade.

Figura 1 – Comparabilidade.



Fonte: Autoria própria.

No exemplo do casamento e da bicicleta, ambos compartilham predicados como custo financeiro, prazer, realização pessoal etc. Esses predicados são os elementos relativos em que é possível comparar os dois portadores. Ainda que

se chegue à conclusão de que os prazeres proporcionados por ambos sejam incomparáveis, eles são o tipo de coisa que podem estar em uma relação comparativa, ainda que se descubra que são pragmaticamente incomparáveis.

Chang (1997, p. 28) considera essa definição importante para que não confundamos o fenômeno da incomparabilidade com o que ela chamou de *não-comparabilidade*, pois enquanto esta é uma falha formal da comparabilidade, aquela é uma falha substantiva. Exemplificando no diagrama, elementos que pertencem apenas ao conjunto A (A - B), que pertençam apenas ao conjunto B (B - A) ou que não pertencem a nenhum dos conjuntos, são elementos em que há uma falha formal na comparabilidade, pois o elemento não é relativo (comum) a ambos os portadores de valor. Podemos comparar uma viagem à Itália a uma xícara de chocolate quente em relação ao prazer produzido, mas não podemos compará-las em relação ao sabor, pois este é um elemento pertencente a apenas um dos portadores comparados.

1.5 COMPARATIVISMO

Um chão comum entre as teorias de justificação prática é o fato de que a opção escolhida, para ser considerada racionalmente justificada, não deve possuir nenhuma razão contrária mais forte do que ela. Isso não implica que as comparações sejam condições necessárias para que haja justificação racional. Como dito acima, o comparativismo é a ideia que considera que a comparação seja uma condição necessária para que haja justificação prática racional.

Cientes dos requisitos da comparabilidade, seguimos para a análise das principais visões de justificação prática defendidas na literatura e suas relações com a comparabilidade.

1.5.1 Otimização e maximização

Os conceitos de otimização e maximização são mais frequentemente utilizados na teoria econômica da decisão (SEN, 1997; 2000). Enquanto a otimização sustenta que a alternativa é justificável se puder ser considerada ao menos tão boa quanto às outras (CHANG, 2015, p. 46), a maximização sustenta que a alternativa pode ser justificada se não puder ser considerada como pior

que as demais (CHANG, 2015, p. 52). Para Chang (1998, p. 1577; 2015, p. 46), a otimização é a visão historicamente dominante do comparativismo.

É possível encontrar uma outra definição frequente para otimização, em que atribui a justificação à escolha da melhor opção possível, mas essa definição é vaga sobre o que poderia atender ao requisito de "melhor opção", assim como é silente sobre a possibilidade da existência de mais de uma opção justificável ou se "melhor" deveria ser entendido como "a única melhor" (HSIEH, 2007, p. 71).

A maior diferença entre essas visões é que enquanto a otimização depende necessariamente da comparação, a maximização é compatível com o fato de as alternativas não poderem ser comparadas. Se o que se busca na maximização é apenas que uma alternativa não possa ser considerada pior do que as demais, quando há uma situação de incomparabilidade, essa condição está satisfeita, pois entre itens incomparáveis, não é possível dizer que uma alternativa é pior e nem que dizer que não é pior. Portanto, conquanto não se possa afirmar que uma alternativa é categoricamente pior do que a outra, o critério da maximização estará atendido.

Um ataque comum contra a otimização é assumir que, em sua visão, tudo o que importa é a maior quantidade de um valor: quanto mais, melhor. Como já explicado acima, as dimensões qualitativas não podem ser resumidas, em todos os casos, às dimensões quantitativas de um mesmo valor. A crítica é equivocada porque, a otimização da escolha não precisa se comprometer em resumir à escolha às dimensões quantitativas, mas pode meramente representá-la em termos quantitativos (CHANG, 2015, p. 46).

Contra a maximização[14], Chang (1998, p. 1584) alega que, se pegarmos todos os casos em que entendemos que uma opção não é pior do que a outra, e excluirmos todos os casos de incomparabilidade, a maximização se resume a uma forma de otimização. Isso porque os casos de "não pior que", não-incomparáveis, seriam equivalentes ao "tão bom quanto" da otimização.

[14] A maioria dos autores utilizam o termo *maximazing*. Chang, a partir de 1998, substitui o termo para *maximalizing* (1998; 2015) para ir ao encontro da expressão *maximal elements* usada por Amartya Sen (1995).

A autora aceita o apelo intuitivo da maximização, afinal, parece correto que, diante de opções em que não temos informações o suficiente, estamos legitimados a escolher aquela que, dentro das nossas limitações, não possa ser considerada como a pior. Contudo, sustenta que essa intuição, embora possa servir como uma política pragmática geral em situações de escolha, não implica que possam justificar racionalmente a escolha (CHANG, 1998, p. 1584).

Peguemos um exemplo dado por Raz (1986, p. 332) de uma pessoa que está em dúvida sobre a escolha de qual carreira deve seguir. Ela está inclinada entre uma carreira musical, como clarinetista, e uma carreira jurídica. Do seu ponto de vista, não fica claro que uma das carreiras é uma opção melhor do que a outra. Tampouco, um incremento positivo em uma das carreiras não parece ser o suficiente para a resolução da questão, de modo que elas também não parecem ser igualmente boas.

Sob a ótica da maximização, a pessoa estaria justificada em escolher qualquer das opções, já que nenhuma pode ser considerada pior do que a outra. Entretanto, quando se opta pela maximização, o que está em jogo na escolha (o elemento relativo) não é mais o valor “qualidade como carreira”. Devido a nossa limitação epistêmica, a situação exige que se mude o elemento relativo para o valor “maximização da qualidade como carreira”. Sendo assim, para Chang, é válido como um recurso prático, mas não enfrenta a demanda justificadora. Apenas assume que é impossível supri-la e adota uma estratégia funcional.

1.5.2 Satisfatização (*satisficing*)

Um dos problemas apontados na otimização é o fato de ela fazer parecer que a razão é exigente demais. Pergunta-se, então, se a justificação racional é o tipo de coisa que exige que devemos sempre “levar a melhor” em todas as situações e, caso não seja possível por conta do contexto, que levemos a opção que “empatou tecnicamente” com a outra. Será que a razão demandaria de nós esse comportamento, sob a pena de recusar à nossa ação a alcunha de racional?

Para a satisfatização, a razão exigiria apenas que a alternativa atinja um critério de qualidade satisfatório, isso é, que a alternativa seja boa o bastante, e

não necessariamente a melhor ou a tão boa quanto às demais. Desse modo, estaríamos legitimados a optar por uma alternativa considerada pior do que uma concorrente.

Michael Slote (1989, p.17-18) afirma que uma pessoa vendendo uma casa estaria justificada em aceitar uma oferta dentro de um intervalo de preços, mesmo que soubesse que há uma oferta de maior valor em jogo. Esse exemplo é problemático por várias razões. Mas antes de tentar analisá-lo pela melhor ótica, é essencial afastar suas características contestáveis.

Primeiramente, se for o caso de a maior oferta estar condicionada a qualquer outro fator, então esse fator talvez esteja sendo levado em consideração. Por exemplo, imagine que a pessoa esteja querendo vender rapidamente a sua casa, pois precisa do dinheiro para aceitar uma oferta de investimento. Se a oferta maior estiver condicionada ao tempo, ou seja, o potencial comprador garante a oferta, mas apenas no mês seguinte, ou de forma parcelada, então esses fatores pesam no balanço de razões do vendedor, pois, para ele, é mais importante ter uma quantia razoável de forma imediata, do que uma oferta maior no futuro.

Outro ponto importante é que a venda de um bem tem como finalidade, geralmente, a obtenção de dinheiro. Dinheiro é o tipo de bem que não só podemos medir quantitativamente, como normalmente segue a regra do “quanto mais, melhor”. O aceite da oferta menor parece irracional porque, olhado isoladamente, o ganho de mais dinheiro implica apenas em maior valor ao sujeito.

Alguém poderia contestar a afirmação anterior, mas para isso, seria necessário provar que há alguma razão contrária em aceitar uma quantia maior de dinheiro, de modo que essa razão enfrentasse a razão para aceitar a oferta maior, que é o próprio valor instrumental do dinheiro. Isso é particularmente curioso porque o dinheiro, em si, não possui valor intrínseco. Uma vez que possui exclusivamente valor instrumental e pode ser utilizado em conformidade com as melhores razões do agente em uma infinita possibilidade de opções, fica difícil conciliar a proposta de Slote com a racionalidade.

Uma forma de conciliar esse exemplo com a racionalidade seria pela visão de que é melhor — talvez por uma questão de justiça — aceitar uma oferta razoável do que uma que esteja acima de um determinado valor, pois esta seria

injusta. Há duas maneiras de compreender essa defesa. Imaginemos o seguinte exemplo.

João quer vender sua casa e recebe três propostas: (A) uma de 400 mil reais, (B) outra de 450 mil reais e uma última (C) de 500 mil reais. Digamos que, para João, é uma questão de justiça que ele sempre aceite uma opção que considere uma oferta razoável. Desse modo, ele escolhe a oferta B, pois é a média entre as três alternativas. Parece uma escolha racional, porém, ela não afasta o comparativismo. Para chegar em um resultado razoável, João precisou comparar as alternativas, a fim de encontrar um equilíbrio que justificasse seu critério de “oferta razoável”, que é, nesse caso, o elemento relativo da situação de escolha (CHANG, 1998, p. 1583).

Outra maneira seria imaginar que João já tinha um intervalo de valores que considerava razoável, digamos, entre 430 e 480 mil reais. Entre as três alternativas, apenas a B atendia ao intervalo absoluto definido por João. Dessa forma, a comparação entre as alternativas não seria necessária, contudo, a referência a um padrão absoluto esvaziaria a satisfatização para uma espécie de absolutização (CHANG, 1998, p. 1583), que será vista a seguir.

1.5.3 Absolutização

A absolutização é uma proposta que reivindica que fatos absolutos fornecem os fundamentos que justificam uma escolha. Ela pode ser usada tanto por aqueles que entendem que a comparabilidade tem uma função privilegiada na justificação, quanto por aqueles que entendem que a comparabilidade não é privilegiada ou sequer cumpre qualquer papel na justificativa.

Michael Stocker (1997, p. 207) oferece o exemplo de alguém que alcançou um nível satisfatório de conforto na vida e lhe é oferecido um emprego em que ganharia um melhor salário. Recusar a oferta, para Stocker, não teria qualquer prejuízo à racionalidade, pois o que justifica a vida ser boa é ela ser como ela é, e não o ela ser melhor ou igualmente boa às alternativas de vida disponíveis ao sujeito. Enquanto a satisfatização é instrumentalmente racional, a teoria de Stocker é racional de forma não-instrumental. Ou seja, a satisfatização, conforme elaborada por Herbert Simon (1955), no paradigma

econômico, é um mecanismo racional quando a busca por alternativas melhores é dificultosa ou impossível para a agente. Já em Stocker, o encontro da alternativa boa o suficiente é racional de forma não-instrumental (SLOTE, 2004).

Assim como o exemplo de Slote sobre a oferta de compra da casa, esse exemplo parece estar dizendo algo de forma implícita. Poderia estar implicando a manutenção de um certo “valor conservativo” da vida. Escolhas que tomamos para o futuro, por mais que um exemplo tente isolar variáveis o máximo possível ("todas as coisas mantidas iguais"), dificilmente são assim na realidade. No exemplo acima, ainda que a única coisa que mude na vida da pessoa seja a empresa em que trabalha, isso em si já implica mudanças como novos colegas de trabalho, novos chefes, novo prédio etc. Escolhas que causam possíveis consequências relevantes tendem a ser preteridas em relação à vida que já se vive, se a vida que já se vive é boa ou confortável o bastante.

Me parece que o apelo intuitivo do exemplo reside no fato de que a vida boa que o sujeito está vivendo é a sua vida como ele conhece, portanto, confortável. A mudança de trabalho acarretaria algum desconforto e incerteza, como para se adequar a uma nova rotina, por exemplo. Talvez o sujeito tenha pensado que é mais valioso manter uma vida boa que se conhece do que se arriscar em uma nova vida que desconhece. Contudo, essa avaliação parece acarretar um cálculo probabilístico do tipo “custo-benefício”, em que são pesadas as razões positivas de um salário maior contra as razões negativas de sair da zona de conforto. Se for esse o caso, existe uma comparação.

Chang (1998, p. 1587) diferencia duas formas de absolutização: as baseadas em mérito e as baseadas em regras. O exemplo de Stocker seria baseado em mérito, uma vez que apela para uma certa qualidade de vida. Considerações baseadas em regras, por sua vez, apelariam a um padrão de comportamento ou outro tipo normativo. Por exemplo, em uma mesa de jantar, alguém estaria justificado em colocar o guardanapo em seu colo, pois se trata de uma regra de etiqueta à mesa. Essa forma parece mais plausível que a baseada em mérito, pois evita o problema da justificação elíptica[15]. Porém, ainda

[15] Chang aponta que em situações de absolutização baseadas em mérito, a conclusão do fundamento parece sempre se fundamentar na premissa de forma elíptica. Por exemplo, dizer que uma determinada ação é justificada porque atende a um certo padrão só pode ser

que a versão baseada em regras possa justificar racionalmente a ação, alega-se que ela não implica na derrota do comparativismo, pois este, em sua forma indireta, explicaria como a referência a um fato absoluto continua dependendo da comparabilidade, conforme se analisará no tópico seguinte.

1.5.4 Comparativismo indireto

Segundo Chang, as visões incomparabilistas (que não concordam com o comparativismo), vistas acima, falham porque assumem como natural uma caraterística que não é um requisito necessário em todas as formas de comparativismo. Essa assunção se resume em inferir que aquilo que justifica a escolha é a mesma coisa que determina a escolha como justificada (CHANG, 1998, p. 1572). Tal inferência pertenceria a uma forma *direta* de comparativismo, que poderia ser contrastada com uma outra forma — o comparativismo *indireto*.

De acordo com a versão indireta de comparativismo, não há uma transição automática entre o que *justifica* a escolha e o que *determina* a escolha como justificada, pois o que quer que determina a escolha vem daquilo que fornece uma *força justificadora* da justificação (CHANG, 1998, p. 1573). Chang não se compromete com uma teoria que derrota a versão direta do comparativismo, mas somente com a possibilidade de que ele exista como gênero, do qual as formas direita e indireta sejam possíveis espécies. A teoria continua sendo comparativista, pois tal força justificadora só pode vir de um elemento comparativo entre as alternativas.

A autora difere a força justificadora do próprio fundamento justificador apelando a uma analogia com o estudo da lógica proposicional. Considerando as inferências lógicas, quando pensamos em uma proposição $p \rightarrow q$ (se p então q), pergunta-se: o que justifica, em tal proposição, a inferência de q das premissas p e $p \rightarrow q$? As premissas sustentam a conclusão, mas em virtude *do que* elas sustentam é a regra de inferência *modus ponens*[16]. Pela analogia, a

compreendida de forma coerente se entendermos que o elemento relativo é fornecido por esse padrão (1998, p. 1587)

[16] Na lógica informal, o mecanismo do *modus ponens* é uma forma de validação das conclusões. Essencialmente, a ideia de validade do argumento existe para evitar o problema da relação contextual entre conclusões e premissas que surge da afirmação que determinada conclusão

regra lógica do *modus ponens* não é parte da conclusão, mas dá suporte, ou "poder lógico" para que produzir a conclusão (CHANG, 1998, p. 1589).

Um segundo exemplo fornecido, apela a uma intuição de ordem física. Se imaginarmos uma vidraça quebrada por um tijolo, podemos dizer que o que causou a quebra da vidraça foi o impacto com o tijolo. Porém, o impacto do tijolo é a causa em virtude de certas leis nomológicas que fornecem ao impacto uma certa "força causal" que acarreta a quebra da janela. Assim como as leis nomológicas não fazem parte daquilo que consideramos a causa da quebra da janela, pois apenas tem o poder de causar, a força justificadora de uma justificação não se confundiria com o próprio fundamento justificador (CHANG, 1998, p. 1589).

Para Chang, há duas formas de explicar a determinação causal da quebra da janela. Na primeira, há um apelo a uma consideração saliente, que seria a causa: o arremesso do tijolo. Essa explicação, em termos de causa, precisa de certas condições anteriores para que a causa explique a quebra da janela. A outra forma de explicar — a não mediada — permite dizer que um conjunto de condições anteriores determina a causalidade da quebra da janela sem apelar a outras condições. Não haveria, nessa forma, uma consideração saliente, mas apenas o estado de coisas detalhado, "talvez dado em um nível atômico" (CHANG, 2015, p. 59).

Os esforços de Chang são para que possamos compreender que há um elemento, talvez similar a uma energia metafísica, que surge necessariamente das comparações e que é esse elemento que produz (ou permite) a racionalidade da justificação da escolha. Assim, um fundamento justificaria a escolha não por algo inerente a si, mas por meio de uma força normativa surgida do fundamento (CHANG, 1998, p. 1589). Um exemplo mais atraente, que busca

---

segue de certas premissas. Assim, um argumento é dito válido se e somente se não for possível que todas as suas premissas sejam verdadeiras e a conclusão seja falsa. O *modus ponens* é uma forma de estruturação de um argumento que busca esse conceito de validade ao demonstrar que, em determinado caso, é impossível que premissas sejam verdadeiras quando a conclusão é falsa. Um exemplo desse mecanismo pode ser observado no seguinte argumento: Premissa 1: "Se está nevando lá fora, então a estrada está escorregadia"; Premissa 2: "Está nevando lá fora" e; Conclusão: "A estrada está escorregadia" (inferência das premissas 1 e 2). No exemplo, se a conclusão for falsa, então é impossível que as premissas se mantenham verdadeiras. (SINNOTT-ARMSTRONG, 2015, p. 92)

enfrentar as alegações das visões incomparabilistas, pede que imaginemos uma situação de promessa entre amigas[17].

Imagine que duas amigas, Ana e Patrícia, estejam conversando e que Patrícia esteja passando por um momento ruim. Ana, no intuito de proporcionar algum prazer à amiga Patrícia, promete que irá animá-la no próximo fim de semana. Ana sente que, ainda que seu fim de semana estava previamente ocupado com seu trabalho de corrigir provas, estará justificada em sair para jantar com Patrícia ao invés, uma vez que tenha feito a promessa. A escolha de sair para jantar, portanto, atende à demanda pela justificação da escolha. Contudo, alega Chang, a saída para jantar só justificou a escolha porque a outra opção disponível era que Ana ficasse em casa corrigindo provas, e esta opção é pior em relação ao conteúdo da promessa (proporcionar diversão à Patrícia) do que aquela (CHANG, 1998, p. 1569).

Ainda que se entenda que sair para jantar seja um fato concreto (absoluto) da alternativa — suficiente para cumprir a promessa — ele só pode justificar a escolha porque houve, antes da escolha, uma situação comparativa que forneceu a força justificadora. Para sustentar essa conclusão, Chang (1998, p. 1569) diz que se tivéssemos à disposição, opções melhores para concorrer com a saída para jantar (como andar de patins na praia), e desde que essas opções cumprissem melhor à promessa de animar a Patrícia, então a opção de sair para jantar não estaria mais justificada.

Hsieh (2007, p. 74) aponta que o comparativismo indireto não é convincente em relação ao descarte da maximização como justificação racional. Ele afirma que o critério de “não ser pior que” possui força justificadora da mesma forma que “ao menos tão bom quanto” (HSIEH, 2007, p. 74). Para elucidar, retornemos à analogia da causalidade no exemplo da janela quebrada. Pela explicação não mediada, em que o tijolo não é saliente, a causalidade se volta para um conjunto detalhado de condições anteriores para determinar a quebra da janela. Dessa forma, ele conclui que:

> Uma vez que a comparação de alternativas não é mais saliente, no entanto, não vejo razão para que uma alternativa que não é pior do que outra não poder ser aquela em virtude da qual a base para a escolha

[17] O exemplo foi ligeiramente modificado do original apresentado pela autora, contudo, preservando a sua essência.

> tem força justificadora. A ausência de uma relação de valor positiva parece não providenciar uma descrição mais completa do que uma relação de valor positiva do estado de coisas que explica a escolha como justificada. Como resultado, uma vez que os proponentes do comparativismo adotam o comparativismo indireto como a versão mais plausível, parece não haver razão para exigir a comparabilidade de alternativas para que o fundamento da escolha tenha força justificadora. (HSIEH, 2007, p. 74).

A dificuldade para aceitar a maximização como teoria da justificação racional parece repousar na ideia do quão flexível é a razão para chancelar a resposta justificada. Enquanto Hsieh tem uma ideia mais ampla do papel da razão na justificação, Chang, conforme já mencionado em 1.5.1, aceita o argumento como válido para uma política pragmática geral, mas recusa a aceitar que isso implica em uma decisão racional, por não atender a demanda por justificação.

Se assumirmos a maximização como teoria da justificação racional, em todas as situações em que há uma escolha difícil, como em dilemas morais, a resposta será dada no sentido de o agente estar legitimado a escolher qualquer das opções. Imagine que alguém esteja em dúvida se deve se mudar de cidade para melhorar de vida, ou se deve continuar na sua cidade para cuidar de seus pais doentes. Uma vez que ele não possa afirmar categoricamente que uma alternativa é pior que a outra, então, ele estaria legitimado a escolher qualquer delas. Contudo, isso parece resolver a questão simplesmente negando o problema. Ainda restaria saber algo a respeito de qual alternativa escolher. Se ambas as opções são racionais, e a razão é indiferente à escolha, então seria legítimo escolher no cara-ou-coroa? Isso não parece ser a atitude correta diante da relevância da escolha. Sendo assim, no próximo capítulo, partiremos da ideia de que a comparação desempenha um papel importante na justificação racional, abordando, entre outras coisas, esse incômodo causado por escolhas feitas entre opções de valores incomensuráveis.

2 A INCOMENSURABILIDADE E A DECISÃO JUSTIFICADA

> *Presents are the best way to show someone how much you care. It's like this tangible thing that you can point to and say, "Hey, man, I love you this many dollars worth".*
>
> Michael G. Scott

O comparativismo parece ter fortes argumentos para sustentar que a comparabilidade é um critério necessário da justificação racional. Como foi visto, algumas visões entendem que a comparabilidade pode ser afastada nos casos em que for impossível, ou seja, nos casos de incomparabilidade entre as opções em relação a um determinado elemento relativo. Além disso, vimos que as visões que entendem que a comparabilidade não possui qualquer função justificadora foram contestadas pelo comparativismo indireto. Assim, iremos assumir, daqui em diante, que as comparações, quando possíveis, são fundamentais para justificar a escolha.

Partindo, portanto, da relevância da comparabilidade para a justificação racional, pergunta-se: o fenômeno da incomensurabilidade oferece algum problema prático para a comparação? Em outras palavras, em uma situação comparativa, a ausência de uma medida comum entre as opções, impede a possibilidade de que se comparem? Se a resposta for sim, então a possibilidade de justificação estará comprometida. Mas, antes de partirmos para o problema em si, o próximo tópico irá definir conceitualmente o que se quer dizer com o termo "incomensurabilidade".

2.1 DEFINIÇÃO TERMINOLÓGICA DE INCOMENSURABILIDADE

O uso do termo incomensurabilidade se divide em dois contextos distintos. O primeiro tem sua relevância na epistemologia e na filosofia da ciência. Nesse contexto, teorias científicas incomensuráveis se referem a paradigmas teóricos que só poderiam ser propriamente entendidos dentro de suas próprias molduras conceituais. Dessa forma, por exemplo, a física aristotélica, que explica o comportamento dos objetos em relação ao seu propósito (*telos*) (CHARLTON,

2006), seria incomensurável com o paradigma da física de Newton, que utiliza o conceito de gravidade e partículas para explicação dos mesmos fenômenos (KUHN, 1996, p. 103-104). Ainda que ambas as teorias possam ser reputadas como válidas de forma independente, só é possível compreendê-las dentro de seu próprio paradigma, não se relacionando conceitualmente uma com a outra. Por isso, são reputadas como incomensuráveis.

O segundo contexto, que será o foco deste trabalho, tem importância para os campos da teoria do valor, teoria normativa e filosofia da razão prática. O problema da incomensurabilidade, nesse contexto, já havia causado algum desconforto intelectual na Grécia Antiga. Antes do conhecimento a respeito dos números irracionais, acreditava-se que era possível representar a realidade em números inteiros, e isso era a base na qual toda a filosofia pitagórica se sustentava (FRITZ, p. 382-412). O filósofo e matemático Hipaso de Metaponto é apontado como o possível responsável pela descoberta desse fato que abalou as crenças da filosofia pitagórica da época. Ao notarem que o comprimento da diagonal de um quadrado não podia ser medido por uma razão de números inteiros, mas sim pela raiz quadrada de dois, os pensadores teriam chamado o fenômeno de *alogos* ou *arrhētos* (ambos normalmente traduzidos como “inexpressível” ou “irracional”). Dada a importância da descoberta, conta a lenda que Hipaso de Metaponto foi afogado pelos deuses por revelar esse segredo (HEATH, 1921, p. 1523). Tempos depois, Aristóteles (1999, 1133b 15–25) se referiu à incomensurabilidade (*asummetros*) de valores como a falta de uma unidade comum pela qual eles pudessem ser medidos.

É possível encontrar o termo incomensurabilidade se referindo a diversos tipos de ideias, sendo que algumas delas serão discutidas adiante. A maioria dessas ideias tem alguma implicação na possibilidade ou impossibilidade da comparabilidade, isto é, enquanto alguns teóricos argumentam que a incomensurabilidade implica necessariamente na incomparabilidade, outros sustentam que esse não é necessariamente o caso. Alguns autores também utilizam os termos incomensurabilidade e incomparabilidade como sinônimos (RAZ, 1986, p. 322). Contudo, ainda que alguém possa sustentar que a

incomparabilidade é decorrência necessária da incomensurabilidade [18], os termos parecem se referir a fenômenos distintos.

Mensurar é um sinônimo para medir. Comensurar[19] significa medir, ou mensurar, dois ou mais itens pela mesma medida. Dessa forma, imensurável é aquilo que não se pode medir e incomensurável se refere a dois ou mais itens que não possuem uma medida comum entre eles. O uso intercambiável dos termos incomensurável e incomparável pode ser por conta de uma confusão linguística.

Há um uso corriqueiro, ao menos no português, do termo "incomensurável" significando algo que é tão grande que seria impraticável medi-lo, mas não necessariamente impossível, como o número de grãos de areia em uma praia. É possível também encontrá-lo significando um elemento que não é necessariamente grande ou pequeno, mas simplesmente impossível de se medir de forma precisa, como valores abstratos do tipo conforto ou tranquilidade. Além disso, também é comum se observar o uso do termo "incomensurável" para se referir a um elemento isolado, que se pretende entender como grandioso, como quando se diz que "justiça é um valor incomensurável" ou "o meu amor por você é incomensurável".

Nos exemplos dados acima, podemos aceitar o sentido figurado do termo. Porém, formalmente, está incorreto por estar ausente o elemento de interseção que justifica a partícula -co. Assim, nos casos corriqueiros como os exemplificados, o termo correto seria "imensurável".

Dadas as raízes etimológicas, este trabalho irá seguir o entendimento dos pensadores que consideram incomensurável como significando a ausência de medida comum entre dois elementos quaisquer (FINNIS, 1980, p. 113; WIGGINS,

[18] Esclareço que esse não o caso do Raz, ou seja, o autor não pensa que a ausência de medida comum implique, necessariamente, na incomparabilidade. De fato, seu argumento do padrão inerente (RAZ, 1991, p. 86-89) sustenta a possibilidade de comparação diante da ausência de medida comum. O que ocorre é que o sentido que queremos avançar aqui como sendo o significado de incomparabilidade é o que Raz chama, ora de incomparabilidade, ora de incomensurabilidade. O sentido que usamos aqui para definir a incomensurabilidade está implícito no que Raz (1991, p. 85) chamou de "argumento comum".

[19] A partícula -CO denota uma conjunção, ou concomitância, como em coautoria, concorrente, correlação etc. Portanto, seu uso é apropriado quando é feita uma referência a dois ou mais elementos em que haja uma certa interseção conceitual, ou concomitância, como em uma relação comparativa.

1987a, 1987b; STOCKER, 1990, p. 175; RICHARDSON, 1994, p. 104; CHANG, 1997, p. 2; SUNSTEIN, 1997; D'AGOSTINO, 2003, p. 35).

2.2 MEDIDA COMUM

A que tipo de coisa estamos nos referindo quando dizemos “medida comum” é um assunto controverso. Para alguns autores, medida comum se refere exclusivamente a escalas cardinais (STOCKER, 1990, p. 176; STOCKER, 1997, p. 203; CHANG, 1997, p. 2). Também chamadas escalas métricas, as escalas cardinais são medidas que indicam quantidades unitárias de valor em um conjunto finito de elementos. Em outras palavras, são aquelas fornecidas por uma régua ou por um termômetro. Por definição, são medidas precisas, universais e independentes da vontade do agente que as utiliza. Por exemplo, a distância entre o Sol e a Terra é de cerca de 150 milhões de quilômetros, independente de quando ou de quem a meça[20]. É possível, assim, comensurá-la com qualquer outra distância que possa ser representada em quilômetros, como a distância entre a Terra e a estrela Markab, na constelação de Pegasus, de modo que podemos dizer não somente qual está mais distante, mas o quanto está mais distante.

É possível encontrar dois tipos de escalas cardinais: intervalares e de razão. As escalas que marcam um zero absoluto (ou definitivo) são as escalas de razão. É o caso de escalas que medem distância (como no exemplo acima), idade, altura, peso etc. Sua característica consequente é a de não possuir valor negativo. Escalas intervalares, por sua vez, não possuem um zero definitivo, de modo que podem ter medições negativas, como algumas escalas de temperatura[21] do tipo Celsius e Fahrenheit. Como estas marcam o intervalo entre os elementos, elas não fornecem uma medida significativa de proporção. Assim, podemos dizer que o ponto de congelamento da água, 32 °F, é menor que seu ponto de ebulição, que é 212 °F, contudo, não podemos dizer que uma porção de água a 100 °F tenha o dobro da temperatura de outra que esteja a 50 °F.

[20] É sabido que existem variações de distância entre os corpos celestes, mas não é necessário esse nível de rigor para fins exemplificativos.
[21] É possível que uma escala de temperatura não seja intervalar, como a escala Kelvin, que possui um zero absoluto.

Alguns pensadores entendem que medida comum deve se referir apenas às medidas representadas por escalas cardinais. Desse modo, devemos entender a incomensurabilidade como a ausência de escalas cardinais entre dois elementos (CHANG, 1997, p. 2; STOCKER, 1990, p. 176; STOCKER, 1997, p. 203). Como não podemos dizer, por exemplo, que ler um livro possui 20 unidades de prazer a mais do que assistir a um filme, não poderíamos comensurar o prazer entre essas duas atividades. Logo, o prazer entre ler um livro e assistir um filme seria incomensurável, nesse sentido.

O conceito de medida comum pode incluir também as escalas ordinais, quando itens são medidos por uma ordem hierárquica do tipo primeiro, segundo etc. Para os fins deste trabalho, o entendimento será de que a incomensurabilidade ocorre quando não for possível medir os itens avaliados em termos de escalas cardinais ou ordinais.

2.3 NOÇÕES DE INCOMENSURABILIDADE

Partindo da ideia de incomensurabilidade como a ausência de medida comum, algumas características conceituais adicionais foram elaboradas na literatura a respeito. Abaixo veremos duas dessas concepções.

2.3.1 *Trumping* e descontinuidade

Alguns filósofos entenderam a incomensurabilidade em termos de restrições da atuação de um valor sobre um outro valor. James Griffin se referiu a esse fenômeno como *trumping* [22] . Assim, qualquer instância de um determinado valor, não importa o quão pequeno, seria sempre superior (triunfaria) a qualquer instância de outro determinado valor, não importa o quão grande (GRIFFIN, 1986, p. 83). Nesse sentido, o valor da visão, por exemplo, poderia ser considerado incomensurável com o valor de apreciar o sabor do café, pois, não importa o quão grande seja a instância de apreciar café, ela nunca superaria o valor de se enxergar, mesmo que em uma instância mínima do valor visão.

[22] Um sentido similar de *trumping* pode ser encontrado em Ronald Dworkin, ao se referir a determinados direitos morais que deveriam triunfar sobre certas ações políticas (1977; 1984).

Griffin apresenta uma forma mais fraca de *trumping*, chamada "descontinuidade" (*discontinuity*). Nessa forma, há um limite na instância dos valores a serem analisados. Desde que tenhamos uma quantidade específica mínima de um valor A, então, qualquer quantidade de um determinado valor B não irá superar quantidades adicionais ao valor A (que estejam acima de seu valor mínimo) (GRIFFIN, 1986, p. 85). Utilizando o exemplo acima, poderíamos indagar se seria possível trocar o prazer de apreciar café pela visão de alguém[23]. Os valores seriam descontínuos se, por exemplo, estipulássemos que qualquer quantidade acima de 50% da visão seria inegociável com o apreciar do café, mesmo que isso incluísse a degustação dos melhores cafés do mundo. Entretanto, instâncias inferiores a 50% poderiam ser superadas por grandes instâncias da apreciação de café. Alguém estaria "permitido" em negociar, por exemplo, 10% de sua visão para obter um paladar de *Q-Grader* com acesso às melhores safras de café disponíveis.

Chang (2013, p. 4) considera que as concepções de *trumping* e de descontinuidade são úteis para caracterizar as teorias éticas deontológicas. Enquanto a ideia de *trumping* seria usada para conceituar paradigmas mais rígidos, no sentido de que deveres sempre triunfariam sobre a utilidade prática de não seguir um dever, a ideia de descontinuidade serviria para caracterizar visões deontológicas mais moderadas. Nessa última forma, se reconheceria o valor privilegiado de um dever até um certo limite da utilidade de agir contrariamente ao mesmo dever. Por exemplo, considerando que temos um dever de não mentir, na hipótese de que a falta com a verdade fosse a única forma de salvar a vida de um amigo, então esse dever poderia ser relativizado e seria permitido mentir[24].

---

[23] É verdade que o exemplo poderia ser mais atraente o se utilizássemos o dinheiro em troca da visão. É razoável imaginar que alguém estaria disposto a vender uma de suas córneas em troca de alguma compensação financeira. Contudo, por conta do fato de que o dinheiro é um bem de valor instrumental, ele depende de como será utilizado para que seja avaliado adequadamente na situação de escolha. Alguém que vendesse uma das córneas para pagar a cirurgia que salvaria sua mãe de uma doença fatal parece diferir de alguém que usasse a mesma quantia para comprar um carro esportivo, por exemplo. Por isso, a preferência de utilizar dois valores intrínsecos.

[24] Um argumento nesse sentido foi desenvolvido por Benjamin Constant, em 1797, alegando que nenhum homem teria direito a uma verdade que prejudicasse outro homem. Kant responde a Constant negando que tal direito pudesse existir.

2.3.2 Não-compensabilidade

Alguns autores defenderam que a incomensurabilidade tem como característica central a não-compensabilidade na realização de um valor em detrimento de outro. Isso significa que valores seriam incomensuráveis sempre que, em uma situação de escolha, o ganho em optar por um valor não cancelasse totalmente a perda de não optar pelo valor preterido (ANDERSON, 1997; SUNSTEIN, 1997). Escolhas com essa característica parecem ser frequentes na vida humana. Imagine que Pedro seja o tipo de pessoa que gosta de ciclismo e de jogos de tabuleiro, e que está em dúvida sobre qual garota deveria tentar namorar. Aline é ciclista, mas não gosta de jogos, e Joana é entusiasta de jogos de tabuleiro, mas não gosta de esportes ao ar livre. Como Pedro só pode tentar perseguir uma das opções, sente que qualquer que seja a sua escolha, haverá um ganho em compatibilidade de interesses, mas haverá também uma perda não compensada por esse mesmo ganho. O exemplo representa, em partes, o que John Rawls (1971, p. 27) chamou de “separação de pessoas”, em que sustentou que os problemas ou o sofrimento de uma pessoa não poderiam ser compensados pelos benefícios trazidos por outra pessoa.

A característica da não-compensabilidade também é entendida como fundamental para explicar os problemas de dilemas morais (RICHARDSON, 1994, p.115-117; NUSSBAUM, 1990; HARRIS, 2006), fraqueza da vontade (WILLIAMS, 1973, p. 175; WIGGINS, 1987b, p. 239; STOCKER, 1990, p. 230) e o arrependimento pela renúncia de uma escolha inferior (ANDERSON, 1995; LUKES, 1997). Segundo Hsieh (2021), o paradigma da não-compensabilidade negaria a tese da comensurabilidade forte defendida por Henry Richardson (1994, p. 104-105). Essa tese sustenta que sempre é possível avaliar um valor em relação a outro valor em termos de um único valor relativo em comum. Isso implica que haveria uma medida comum de valor universal para todos os bens, de modo que os valores sempre pudessem ser resumidos a instâncias de um supervalor. Negar essa medida comum, contudo, não afastaria a versão fraca de Richardson (1994, p. 105) da tese da comensurabilidade, que sustenta que, em

todos os conflitos de valor, sempre haverá algum elemento possível que possa comensurá-los.

2.3.3 Incomensurabilidade fraca e incomensurabilidade forte

De forma similar às teses da comensurabilidade, as teses da incomensurabilidade também podem vir nas formas forte ou fraca. A incomensurabilidade fraca alega que não existe uma medida comum singular que possa comensurar todos os pares de valores em conflito. Já a tese da incomensurabilidade forte sustenta que, em qualquer combinação de valores em conflito, não será possível obter uma medida comum entre eles (WIGGINS, 1987b, p. 259; RICHARDSON, 1994, p. 104-105). Se a tese fraca estiver correta, então os cálculos de custo-benefício utilizados em teorias econômicas não seriam uma forma confiável de tomada de decisão. Nesse sentido, Aristóteles (1999, 1104b 50–5a1), ao negar que o dinheiro pudesse ser uma medida para todos os valores, pode ser entendido como um defensor da incomensurabilidade fraca.

A incomensurabilidade fraca também é utilizada para sustentar a alegação de que alguns valores possuem um *status* superior a outros, ou são sagrados, de modo que não podem sequer figurar em uma relação comparativa com valores “ordinários” ou de *status* inferior, como o dinheiro, por exemplo (ANDERSON, 1995, p. 141-210; LUKES, 1997, p.187-188). Assim, nos parece contraintuitivo que todos os valores possam ser medidos por uma única medida comum. Se pudéssemos, por exemplo, comensurar o valor da vida de um ente querido com o valor da leitura de bom romance, por meio de, digamos, unidades de um valor, então, encararíamos as perdas desses valores por um prisma meramente quantitativo. Isso sugeriria, talvez, que a morte de alguém pudesse ser compensada pela leitura de muitos livros, de modo a “preencher” a quantidade de valor que foi perdida, mas isso soa altamente implausível. No mesmo sentido, as teorias utilitaristas tradicionais que privilegiam um valor como medida universal, como o prazer, em Jeremy Bentham (2000), também estariam ameaçadas pela tese da incomensurabilidade fraca.

As noções de incomensurabilidade expostas neste tópico servem para ilustrar um pouco da extensão do problema. Para fins de conceituação, contudo, não parecem ser adequadas. As noções de *trumping* e de descontinuidade não apresentam critérios necessários para uma definição do conceito, pois nem todos os problemas de ausência de medida comum são os problemas apresentados por esses fenômenos. Quanto a não-compensabilidade, ainda que se defenda que é uma consequência necessária da incomensurabilidade, parece se referir a um fenômeno distinto, da mesma forma que a incomparabilidade, que se verá no próximo tópico.

2.4 INCOMENSURABILIDADE IMPLICA INCOMPARABILIDADE?

Fica claro que no caso de itens comensuráveis, a comparabilidade é possível. Podemos até dizer que são essencialmente casos de comparações fáceis. Por exemplo, se utilizando uma escala de pés, podemos aferir que Bruno tem 5 '9 pés de altura e Pamela tem 4'11 pés, podemos compará-los em relação à altura e concluir que Bruno é mais alto. Assim, é notório que se há comensurabilidade, então haverá comparabilidade. O problema aparece quando as comparações ocorrem entre itens incomensuráveis — situações as quais irei me referir como *casos difíceis*. Embora soe intuitivo que negar a comensurabilidade implicaria em negar a comparabilidade, esse não é um raciocínio que decorre de uma consequência lógica necessária. A pergunta que se faz, portanto, é se é possível haver comparabilidade entre incomensuráveis.

Sobre a *dificuldade* das comparações, é importante fazer uma diferenciação conceitual. A dificuldade pode ser compreendida de duas formas. A primeira é a dificuldade em resolver substancialmente o problema. Ou seja, apresentar uma resposta verdadeira sobre como os dois elementos se relacionam em relação a algum valor. A segunda dificuldade tem relação com o próprio ato de escolher. Nem sempre comparações difíceis são importantes o suficiente para serem qualificadas como escolhas difíceis. Se eu estou em dúvida entre dois conjuntos de xícaras para ornar com a prataria da minha cozinha e não consigo decidir a questão, pois é difícil extrair uma resposta verdadeira sobre como esses dois conjuntos se relacionam em relação ao *design*,

essa não parece ser uma escolha que leva ao sofrimento do escolhedor. Claro que algumas pessoas podem dar mais importância a esse tipo de escolha do que outras. Pode ser que a escolha de um conjunto de xícaras qualifique alguém para a final de um campeonato de design de interiores, de modo que a dificuldade em escolher seja subjetiva e dependente do contexto. Mas o importante aqui é apenas esclarecer que, quando utilizarmos os termos caso difícil, comparação difícil, ou até mesmo escolha difícil, o que estamos nos referindo é à dificuldade em extrair uma resposta verdadeira, e não necessariamente a dificuldade do ato de escolher.

Algumas respostas podem ser dadas a respeito da possibilidade de haver comparabilidade em casos difíceis. A primeira delas é a mais direta e afirma que não é possível comparar itens incomensuráveis. Por essa visão, a medida comum é um critério indispensável para que possamos comparar os elementos. Portanto, havendo incomensurabilidade, haveria consequente incomparabilidade. Chamarei essa visão de incomparabilista. A segunda resposta aceita a comparabilidade entre incomensuráveis, porém acrescenta que, devido à limitação epistêmica humana, não é possível que conheçamos todas as variáveis que interferem na comparação, de modo que nossa ignorância impede a avaliação do resultado. Essa visão chamarei de epistêmica. A terceira resposta sustenta que nossas definições a respeito do que são relações de valor positivo são conceitualmente vagas. Por isso, ao menos algumas comparações entre incomensuráveis são casos semânticos limítrofes. Em outras palavras, aquilo que entendemos como “melhor”, por exemplo, é um conceito vago e que, embora possa ser descoberto em algumas comparações, em outras poderá ser indeterminado, de modo que a resposta não poderá ser definida nem como sendo melhor, nem como pior (ou igualmente boa). Essa visão chamarei de indeterminismo semântico.

Há um evidente problema prático nessas três respostas: elas não dizem nada a respeito de como se deve decidir. Seja a natureza do problema a ignorância, a incomparabilidade ou a indeterminação dos predicados, e partindo do pressuposto de que, ao menos em algumas situações, não é possível que o agente se abstenha da decisão, nos resta saber se há alguma ferramenta que fundamente racionalmente a escolha. Se a razão não desempenhar qualquer

papel nessas situações, então nos resta apenas a conclusão de que nenhuma opção é privilegiada e, portanto, decidir a escolha no “cara-ou-coroa”, por exemplo, cumpriria ao menos o propósito prático (a exigência da decisão). Porém, o desconforto causado pela escolha aleatória nos indica que ela pode não ser suficiente, especialmente nos casos em que a escolha não seja apenas difícil, mas tenha grande relevância.

Para resolver o problema da escolha, Chang enfrenta as três diferentes visões sobre os casos difíceis, oferecendo uma teoria que não apenas resolveria o problema prático acerca de como decidir, como reivindica ser o que é exigido pela racionalidade. Essa teoria foi mencionada no capítulo anterior para definir conceitualmente os requisitos da comparabilidade[25], e será vista adiante com maior profundidade.

2.4.1 A paridade como solução dos casos difíceis

As possíveis relações comparativas de valor entre dois itens foram tradicionalmente compreendidas pelas três opções: melhor que, pior que e igualmente boa. Esse limite, muitas vezes assumido como um truísmo pelos teóricos, representa o que Chang (2002, p. 660) chamou de tese da tricotomia. Sua teoria da paridade busca superar a tese da tricotomia, ao avançar o argumento de que existe espaço conceitual para uma quarta relação de valor positivo entre os itens em situação comparativa. Isso significa que, caso os itens não estejam relacionados por uma das três relações tricotômicas, então eles podem estar em paridade[26].

A defesa da paridade se inicia com o apelo intuitivo de que existem situações em que conseguimos enxergar a comparabilidade, embora não seja possível extrair uma relação tricotômica. Se entre dois itens, não é possível

[25] Ver 1.4.2.

[26] Apenas como curiosidade, é interessante que Chang tenha escolhido o termo paridade para nomear a sua teoria. Isso porque ela reservou o termo incomensurabilidade à ausência de escala cardinal por conta de um apego etimológico do termo e, além disso, ao analisar a comparabilidade entre incomensuráveis, despreza o apego matemático de entender a comparabilidade de forma precisa (ou exata). Ou seja, ao avançar uma teoria da comparabilidade imprecisa, ela se filia à corrente de que a comparação não requer rigor matemático. Contudo, o termo paridade se remete etimologicamente à matemática. Paridade é a condição daquilo que é par, de números que podemos dividir em partes iguais sem que se sobre um resto. (HUGO DE SÃO VÍTOR, 2020, p. 188)

encontrar uma resposta em termos de melhor que, pior que e igualmente bom, no entanto, é possível compará-los, há uma indicação de que uma quarta relação de valor esteja presente. Sendo assim, antes de buscarmos a quarta relação, precisamos nos certificar de que, de fato, a relação existente na comparação observada não é uma das três relações padronizadas.

A investigação que busca qual relação existe entre os itens a serem comparados propõe um exercício hipotético. Conforme foi mencionado no exemplo das carreiras de Raz [27], se pudéssemos conceber um pequeno incremento de valor em uma das opções que não fosse o suficiente para encontrar uma relação de valor positivo em termos de melhor que, pior que e igualmente bom, então, estaríamos diante de uma escolha difícil. Esse exercício, chamado de argumento da pequena melhoria (CHANG, 2002, p. 667), foi conceituado por Chang da seguinte forma:

> Se (1) A não for nem melhor nem pior do que B (em relação a V), (2) A+ for melhor do que A (em relação a V), (3) A+ não for melhor do que B (em relação a V), então (4) A e B não estão relacionados por nenhuma tricotomia padrão de relações (relativizado para V). (CHANG, 2002, p. 667).

O interessante a respeito dessa ferramenta é que ela exclui as opções tricotômicas, mas deixa espaço para uma quarta relação de valor. Afinal, ela apenas demonstra que não é possível que os itens se relacionem por uma das três relações tradicionais, o que não implica na exclusão de uma outra relação, seja ela qual for. Partindo da ideia de que existem situações em que é possível ter um acréscimo (ou decréscimo) em um dos itens, sem que isso implique na conclusão de que um é melhor do que outro, ou que são iguais, Chang avança sua teoria da paridade.

A defesa principal feita pela autora se apresenta pelo que ela conceituou como argumento do encadeamento unidimensional (CHANG, 2002, p. 673). Partindo de uma situação em que se aplicou o argumento de pequena melhoria, o argumento pede que imaginemos um *continuum* que se inicia com a tentativa de comparar dois itens quaisquer. Em seu exemplo, Chang sugere que tentemos comparar Mozart e Michelangelo em relação à criatividade. Em seguida, devemos imaginar uma sequência contínua de escultores que são levemente

[27] Ver 1.5.1.

piores que Michelangelo até que tenhamos Talentlessi, um escultor particularmente ruim. Talentlessi é um sósia quase perfeito de Michelangelo, com exceção apenas de um único valor prejudicado. Se podemos comparar Talentlessi à Mozart, então podemos comparar Michelangelo à Mozart, pois, de acordo com a intuição, acréscimos ou decréscimos unidimensionais em um único aspecto de um portador de valores não pode desencadear incomparabilidade onde havia comparabilidade (CHANG, 2002, p. 673). Em outras palavras, se em um extremo do *continuum* é possível verificar a comparabilidade, não é justificável que em um ponto intermediário haja incomparabilidade se houve apenas uma gradação de uma característica unidimensional de um dos portadores.

Embora haja um forte apelo intuitivo no argumento do *continuum*, ele não é universalizável, isto é, não pode ser aplicado por extensão a todas as situações comparativas. O que Chang busca, de fato, não é uma regra geral para a comparação em casos difíceis, mas somente a abertura do terreno exploratório em busca de uma quarta relação de valor. Por isso, admite que, em certos casos, o apelo ao *continuum* não é adequado. Ele será válido apenas nas situações em que a dificuldade de encontrar uma resposta avaliativa resida na forma em que podemos equilibrar a qualidade dos valores em jogo. Por exemplo, em uma opção entre carreiras, a dificuldade em escolher pode estar no fato de que uma é relevante no aspecto financeiro, enquanto a outra é relevante no aspecto de satisfação pessoal. Pela ausência de um mecanismo que defina como esses dois valores (dinheiro e satisfação) se relacionam, a escolha se torna difícil. Mas nem sempre esse é o caso.

Algumas situações em que a solução não depende do equilíbrio entre os valores são aquelas em que podemos aplicar uma regra de Pareto. Pensemos no bem-estar. O bem-estar entre duas pessoas não possui uma medida comum (ao menos não uma medida precisa), de modo que a comparação entre seus estados parece ser do tipo difícil. Contudo, se duas pessoas estão em estado de desigualdade de recursos, digamos, em relação a quantidade de alimentos que possuem, e A possui mais suprimentos do que B, então, se aumentarmos a quantidade de alimentos de B (sem retirar de A), esse novo estado de coisas, em relação ao bem-estar social, será superior ao anterior. Isso ocorre porque,

pelo paradigma da superioridade “paretária” (que usam a regra de Pareto), uma situação será melhor se houver um acréscimo de qualidade em um aspecto e um decréscimo em nenhum aspecto (RAZ; GRIFFIN, 1991, p. 87; CHANG, 2002, p. 676).

Em relação às três respostas à incomensurabilidade, dadas no início deste capítulo, Chang afirma que cada uma possui uma virtude, mas compartilham do mesmo vício. No caso da visão incomparabilista, a virtude está em assumir que nem sempre os itens se relacionam por uma das relações tricotômicas. No caso da visão epistêmica, há precisão na assunção de que existe a comparabilidade, mesmo entre incomensuráveis. Já a visão indeterminista acerta na suspeita de que existe uma falha no paradigma comparativo. Contudo, segundo a autora, o vício está no fato de que essas visões chegam às suas conclusões comprometidas com a tese da tricotonomia (2002, p. 661), o que as impede de compreender o panorama geral das situações comparativas. Os incomparabilistas, ao verificar que não há uma das três relações, concluem pela incomparabilidade; os epistêmicos, concluem pela limitação cognitiva; e os indeterministas, concluem que as relações tricotômicas comportam zonas de indeterminação semântica.

Se assumirmos a legitimidade do argumento do *continuum*, então parece que as visões incomparabilistas e epistêmicas deixaram de considerar uma quarta relação, o que colocou em xeque suas conclusões. Contudo, o argumento da vagueza não pode ser afastado com tanta facilidade, afinal, a paridade poderia ser apenas um termo para definir os casos em que as relações de valor tricotômicas são indeterminadas.

Em síntese, um conceito é vago se ele possui casos semânticos limítrofes. Por exemplo, o adjetivo “alto” é vago, pois é possível que existam pessoas que não podemos definir de forma categórica se são altas ou não. Em outras palavras, não há uma medida exata que diga que, a partir dela, alguém pode ser considerado alto. Ao invés, existe uma região indeterminada em que as pessoas podem ser consideradas altas ou não. Essa região é chamada de zona de indeterminação, e as instâncias que se encontram dentro dela são chamadas de casos limítrofes. Tais casos podem ser relativos ou absolutos. Nos casos relativos, a questão é clara, mas os meios de decidi-la não são; enquanto nos

casos absolutos, a própria questão carece de completude (SORENSEN, 2018). O fenômeno da vagueza é o cerne do Paradoxo de Sorites[28].

O Dicionário de Filosofia e Psicologia apresenta a seguinte definição para a vagueza:

> Uma proposição é vaga quando há estados possíveis de coisas a respeito dos quais é intrinsecamente incerto se, caso tivessem sido contemplados pelo orador, ele os teria considerado como excluídos ou permitidos pela proposição. Por intrinsecamente incerto, não se quer dizer incerto por conta de qualquer ignorância do intérprete, mas porque os hábitos de linguagem do orador eram indeterminados. (PEIRCE, 1902, p. 748).

Sendo assim, o que foi defendido até o momento para sustentar a existência de uma quarta relação de valor, poderia ser compreendido apenas como um problema inerente a uma falha da linguagem. Assim, saberíamos claramente quando um valor é melhor ou pior do que o outro, da mesma forma que sabemos que o Grigori Rasputin era claramente um homem alto, e Ronnie James Dio era claramente baixo, mas teríamos dúvida sobre a caracterização de “alto” quando essa clareza começasse a desaparecer, como no caso de um homem de 1,85 m de altura[29].

O argumento do *continuum* possui uma semelhança estrutural com os casos de indeterminação. Para demonstrar, peguemos um dos exemplos clássicos do Paradoxo de Sorites, o caso do careca. “Careca” é um conceito vago, pois comporta casos de indeterminação, em que alguém pode ser considerado tanto careca como não-careca. A única coisa que faz com que alguém seja careca ou cabeludo é a quantidade de cabelos que possui na cabeça, porém, essa quantidade é indeterminada. Se pegarmos como exemplo uma pessoa que possui zero fios de cabelo em sua cabeça e acrescentamos um único fio, ela não deixará de ser careca. Sendo assim, a adição de um único fio

[28] *Sorites*, em grego, significa monte, pilha etc. O paradoxo de Sorites pode ser caracterizado pela indagação de quantos grãos de areia é necessário remover para que um monte deixe de ser considerado como um monte. O paradoxo em si ocorre pelo seguinte raciocínio: 1. Um grão de areia não é um monte; 2. Um milhão de grãos de areia são um monte; 3. Se um grão de areia não é um monte, então 1 grão + 1 grão também não são; 4. Se n grãos de areia não são um monte, então n+1 também não são; 3. Se 99.999 grãos de areia não são um monte, então 1 milhão de grãos também não são. (HYDE; RAFFMAN, 2018)

[29] Uma pessoa com essa altura seria considerada claramente alta entre os Mbutis (pigmeus do Congo) e claramente baixa entre os Massais (indígenas do Quênia). Mas, para o bem do exemplo, consideremos que, no geral, alguém com essa altura pode ser considerada tanto como uma pessoa alta, como uma pessoa não-alta (não necessariamente baixa).

de cabelo a um careca não é determinante para que ele deixe de ser careca. Contudo, se continuarmos adicionando fios de cabelo ao careca, em algum momento ele deixará de ser careca e será cabeludo, mas não sabemos ao certo qual fio de cabelo desencadeou a mudança do predicado. Ou seja, ao construirmos um *continuum* em que, a cada instância, um fio é adicionado ao careca, em um ponto da sequência teremos alguém com tantos fios de cabelo na cabeça que será claramente cabeludo. Esse exemplo é muito similar ao *continuum* de Mozart e Michelangelo.

Sobre essa similaridade estrutural, Chang (2002, p. 680) argumenta que a mera semelhança de forma não implica que o *continuum* seja um tipo de sorites. Para isso, ela apresenta um exemplo de argumento que, embora seja semelhante à estrutura de um sorites, não possui uma zona de indeterminação. Ela diz que "(...) se N é um número natural, então também será natural $N + 1$; portanto, para cada $N = 0, 1, 2, 3, \ldots, N$ será um número natural" (CHANG, 2002, p. 680). Contudo, tal exemplo parece não ser suficiente, na medida em que, se dispuséssemos tal argumento em forma de *continuum*, todas as instâncias seriam prontamente reconhecidas como determinadas, ou seja, como números naturais. Não há sequer vestígio intuitivo de vagueza nesse exemplo. Ao contrário, o exemplo que compara Michelangelo e Talentlessi, assim como o exemplo do careca, possuem um intervalo entre as instâncias em que o julgamento avaliativo é indeterminado — não é possível determinar se o homem é careca, e não é possível determinar se Mozart é melhor, pior ou igualmente criativo a Michelangelo.

Além disso, pode ser o caso de que o argumento de pequena melhoria seja dependente da vagueza das relações tricotômicas. Assim, os casos considerados difíceis seriam apenas instâncias contidas na zona de indeterminação de um dos três comparativos "melhor que", "pior que" e "igualmente bom". Chang oferece, essencialmente, dois argumentos para demonstrar que não é o caso. Ambos são argumentos excludentes, ou seja, tentam isolar características distintas entre casos difíceis e casos limítrofes para diferenciá-los conceitualmente. Em outras palavras, a autora identifica particularidades que existem em um caso e não existem no outro, implicando

que tais diferenças fazem parte da natureza de cada um, de modo que não podem ser confundidos entre si.

O primeiro argumento faz referência à fenomenologia dos casos. Nos casos limítrofes, nossa percepção nos leva a crer que qualquer uma das opções disponíveis se aplica à instância em análise. Por exemplo, a beleza pode ser considerada um conceito vago e, portanto, contendo uma zona de indeterminação. Suponha que um grupo de pessoas estejam debatendo se a atriz Hilary Swank é bonita ou não[30]. Algumas pessoas podem defender que ela seja bonita, enquanto outras não, mas nenhuma é capaz de oferecer um ponto conclusivo para solucionar o embate. Já os casos difíceis não funcionam dessa maneira. Quando encaramos uma situação de escolha incomensurável, como a escolha das carreiras, não nos parece que uma das opções é, ao mesmo tempo, melhor e/ou pior, mas sim que não é nem melhor e nem pior.

O segundo argumento se refere a forma como solucionamos a questão. Nos casos limítrofes, além da solução óbvia de dizer que é indeterminado, quando é necessário um mecanismo prático para escolher, tudo indica que a escolha arbitrária seja suficiente. Por exemplo, a obesidade pode ser um conceito vago, no sentido de que pode existir uma mesma pessoa que possa ser considerada como obesa e não-obesa. Dada a necessidade de se colher parâmetros mensuráveis para diretrizes de saúde, políticas públicas, prioridades de atendimento etc., a questão pode ser resolvida pela fórmula arbitrária do Índice de Massa Corporal (IMC). Tal fórmula leva em conta apenas os valores de altura e peso do paciente, dividindo seu peso pela sua altura ao quadrado. Assim, se o resultado do cálculo for maior ou igual a 30, o sujeito é considerado obeso. Contudo, o cálculo desconsidera outras variáveis relevantes que influenciam o próprio conceito de obesidade, como a densidade óssea e a proporção entre tecido adiposo e massa muscular. Isso faz com que algumas instâncias, como atletas de fisiculturismo, claramente não-obesos, sejam considerados obesos[31].

[30] Exemplo levemente modificado do episódio “Prince Family Paper”, da série de comédia americana “The Office”.
[31] O pugilista Rocky Balboa seria considerado com obesidade grau 1 em Rocky IV, quando enfrentou Ivan Drago.

No entanto, a estipulação arbitrária não é igualmente permissível quando aplicada aos casos difíceis. Chang (2002, p. 682-685) aponta que há uma "divergência substantiva" nesses casos e, portanto, a arbitrariedade não seria suficiente para resolver o conflito. Essa divergência ocorre porque nos casos limítrofes, não há um "resíduo resolutivo"[32], de modo que a resposta arbitrária encerra a discussão. Em outros termos, se a indeterminação semântica nos leva a uma bifurcação em que nenhum caminho se mostra superior, então a criação de uma saliência pela arbitrariedade é bastante para a solução (amplamente compreendida) do conflito. Já nos casos difíceis, esse critério não tem a mesma força. No caso das carreiras, se fosse determinado de forma arbitrária que a carreira jurídica é moralmente superior, por exemplo, isso não seria suficiente para resolver a questão da escolha.

Ao final, devemos ter uma maneira de escolher entre casos difíceis. Então, como escolher diante de itens em paridade? A solução apontada por Chang (2013, p. 75) pressupõe a teoria que chamou de voluntarismo híbrido. Para a autora, a separação tradicional das razões para agir pelo seu conteúdo é importante, mas, em algumas situações, devemos considerar uma outra classificação: aquela em relação às fontes. Como agentes racionais, temos à nossa disposição algumas razões que são externas e *fornecidas* a nós, e outras razões baseadas na vontade, que são *criadas* por nós. Diante de situações em que os itens se encontram em paridade, é possível que o agente decida de forma impulsiva (*drift*) entre as opções, como decidindo aleatoriamente. Contudo, as razões internas fornecem uma ferramenta mais adequada: o comprometimento (CHANG, 2017, p. 16-19). Ao nos comprometermos com uma determinada opção, por exemplo, com a carreira musical, criamos razões para escolhê-la e para escolher as demais opções relativas ao nosso próprio poder normativo, ou seja, a capacidade de criar razões para nós mesmos.

2.4.2 Críticas à paridade

Ainda que bem elaborados, os argumentos em defesa de uma quarta relação de valor não estão isentos de críticas. Poderíamos nos questionar por

[32] No original, *resolutional remainder.*

que apenas as comparações entre valores teriam o benefício de uma quarta relação comparativa, enquanto comparações que não envolvam valores não possuem. Além disso, é argumentado que aceitar a possibilidade de uma quarta relação de valor implicaria em rejeitar o axioma da transitividade, uma máxima considerada como fundamental para a racionalidade (HSIEH, 2007, p. 66). Diz o axioma que

> (...) para quaisquer três alternativas A, B e C, se, todas as coisas consideradas, A for pelo menos tão boa quanto B e, todas as coisas consideradas, B for pelo menos tão boa quanto C, então, todas as coisas consideradas, A é pelo menos tão boa quanto C. (TEMKIN, 2000, p. 266).

Em relação a isso, Chang (2002, p. 675) dispensa rapidamente a crítica ao dizer que essa assunção é controversa e temos pouca razão para acreditar que a comparabilidade é transitiva. A autora afirma que a força intuitiva da alegação existencial é forte o suficiente para demonstrar a existência de uma quarta relação de valor. A alegação é a de que podemos demonstrar, pelo menos em algumas relações entre um certo X e um certo Y, que há um *continuum* de pequenas diferenças unidimensionais que conecta X a algum Xn, que é claramente comparável com X e claramente comparável com Y.

O problema é que tal alegação, implícita no argumento do encadeamento, não parece evidenciar perfeitamente a comparação entre Xn e Y. Ou seja, no exemplo de Chang, não é evidente que Talentlessi seja comparável ao Michelangelo em relação à criatividade. Se pensarmos que criatividade é um portador de valores, ou seja, um "pacote" que comporta outros valores, na ausência de um argumento convincente sobre a forma como esses valores "internos" se relacionam entre si para representar o que se quer dizer com "criatividade", então, não é autoevidente que o decréscimo em um elemento da criatividade seja suficiente para a conclusão de que Talentlessi seja fatalmente pior que Michelangelo em relação à criatividade.

Outra forma de encararmos o problema é imaginando que criatividade para um músico, como Mozart, é um pacote com valores que nem sempre coincidem com os valores do pacote de criatividade, no tocante a um artista plástico, como Michelangelo. Focando apenas no lado escultor de

Michelangelo[33], chamaremos a criatividade que se refere a escultores como criatividade de modelagem, e a criatividade que se refere a músicos como criatividade musical.

Recorrendo novamente à analogia com a teoria matemática dos conjuntos, é razoável imaginarmos que o conjunto A (criatividade musical) possui o elemento *condução rítmica*. Tal elemento influencia diretamente a percepção que temos da criatividade do músico. Embora alterações de magnitude em relação a esse elemento façam com o que um músico seja melhor ou pior em relação a criatividade e, portanto, comparável nesse aspecto a outro músico, isso não faz diferença para a comparação com um escultor. O conjunto B (criatividade de modelagem) não parece conter o elemento condução rítmica, pois se trata de um tipo de criatividade diferente. Pode ser que chamemos genericamente ambos os conjuntos de criatividade por conta de uma limitação léxica[34] ou porque ambos possuem uma grande parte de elementos em comum, de modo que, para fins de conceituação, é inteligível nos referirmos a ambos pelo mesmo termo.

Se ao invés de usarmos Michelangelo como exemplo, fizermos decréscimos em Mozart em relação a sua condução rítmica, até que cheguemos em Mozarthritis, um músico excepcionalmente criativo, porém com um problema nas mãos que impede o desenvolvimento de sua condução rítmica (e que afeta negativamente a percepção que temos de sua criatividade musical), podemos compará-lo a Michelangelo em relação à criatividade? Aqui notamos claramente o apelo intuitivo de assumir que Talentlessi é comparável ao Mozart. Talentlessi é péssimo em relação a uma característica relativa à criatividade, enquanto Mozart é péssimo em relação a nenhuma. Contudo, se a característica em que se é péssimo estiver presente em um dos portadores, mas não estiver presente em outro, fica evidente que há uma falha formal na comparabilidade, portanto,

[33] Afinal, Michelangelo também era pintor, poeta e arquiteto. Todas essas pressupõem algum tipo de criatividade.

[34] Há plausibilidade nessa alegação se imaginarmos a quantidade de exemplos em que conceitos idênticos podem assumir conteúdos distintos. Quando usamos o conceito confortável para descrever uma poltrona, certamente não é no mesmo sentido que quando usamos para descrever uma comida caseira (*comfort food*). Poderia ser argumentado que se trata de ambiguidade, como no caso de banco (que pode ser da praça ou um de investimentos), mas quando nos referimos a confortável, estamos apelando para um sentimento que, em alguma medida, está presente tanto na comida, quanto na poltrona, ainda que não tenhamos clareza a respeito dos elementos. No caso de banco (e outras palavras ambíguas) fica claro que o sentido está direcionado a objetos ou fenômenos totalmente distintos.

ocorrendo o fenômeno da não-comparabilidade[35]. Quando buscamos um julgamento avaliativo entre Mozart e Talentlessi, pode ser que o que "salte aos olhos" seja apenas o fato de que este é notoriamente ruim em um aspecto da criatividade e aquele é notoriamente bom em todos os aspectos da criatividade. Esse destaque poderia estar nos confundindo em relação a ideia de que sejam comparáveis.

Se for esse o caso, então, o que há entre Mozart e Michelangelo é a não-comparabilidade, pois nem todos os elementos são comuns aos dois. Para deixar mais clara a intuição que quero apontar aqui, deixarei o exemplo um pouco mais "absurdo". Se pedirmos para que um grupo de pessoas compare um carro novo do tipo SUV, como uma Porsche Cayenne, a uma faca *santoku* velha (e quase sem fio) em relação a sua utilidade, é provável que haja unanimidade entre as pessoas de que o SUV seja mais útil que a faca. Mas, caso haja alguém mais observador entre os avaliadores, ele pode fazer a seguinte pergunta: "que tipo de utilidade?". É claro que um SUV é útil para muitas coisas como transportar pessoas e cargas e, no geral, mais pessoas optariam, em uma situação de escolha livre de contexto[36], pelo SUV ao invés da faca, até mesmo pelo valor econômico. Mas se o pacote de utilidade a que nos referimos seja uma utilidade culinária, uma faca, ainda que velha, tem maior serventia para cortar alimentos do que qualquer veículo teria. Da mesma forma, se estivéssemos em busca de uma utilidade de transporte, a faca seria inútil.

2.4.3 A subjetividade das respostas à incomensurabilidade

Foram apresentadas algumas propostas de justificação entre itens incomensuráveis. Alguns argumentos parecem demonstrar características interessantes a respeito da agência humana, contudo, ainda que se adequem bem aos problemas práticos, a dependência a recursos subjetivos — como a

[35] Ver 1.4.3.

[36] Quando digo livre de contexto, isso deve ser entendido de forma ampla. É impossível conceber uma escolha sem contexto algum, mas nesse caso, me refiro a ausência de qualquer fator extraordinário que crie razões fortes em favor de uma das opções. Por exemplo, se alguém, em face da escolha entre uma Porsche e uma faca velha, poderia escolher a última, caso fosse uma peça de colecionador que estime muito, ou seja uma herança de família de valor sentimental inestimável.

vontade e o comprometimento — parecem limitar essas teorias às decisões tomadas por seres humanos. Isso porque não é claro como uma máquina poderia ter ou conhecer preferências humanas, especialmente por conta da natureza dinâmica e individual das vontades. Em outras palavras, é parte da natureza das preferências humanas que sejam temporais (ou dinâmicas) e dependentes do agente (subjetivas). Duas pessoas diferentes podem, diante de uma mesma escolha, tomar decisões racionais operadas pelas suas preferências individuais, chegando, cada uma delas, a uma opção diferente. Além disso, infinitas variáveis podem surgir com o tempo, alterando as preferências dos agentes, de modo que possam optar por uma alternativa diferente — e até contraditória — daquela que optaram no passado.

Se buscarmos parâmetros para decisões tomadas por algoritmos de inteligência artificial, algumas perguntas podem surgir, como a indagação de se o dinamismo e a individualidade das vontades humanas podem ser compreendidas pela tecnologia do *deep learning*. Ou então, caso negativo, se há algum parâmetro justificável para guiar as decisões entre incomensuráveis que substitua tais características das preferências humanas.

No próximo capítulo, irá se discutir brevemente o propósito da inteligência artificial para que, com base nessa premissa, possamos identificar o que devemos esperar dela. Em seguida, será feita uma análise das características peculiares das preferências humanas, buscando um caminho para a sua definição e, posteriormente, para o desafio da escolha.

3 A INTELIGÊNCIA ARTIFICIAL E AS ESCOLHAS DIFÍCEIS

> *O cérebro humano é um gênio, porém, muito lento. O computador é um idiota, porém, muito rápido.*
>
> Prof. Pierluigi Piazzi

Costuma ser um grande desafio fazer a definição de um termo, e com a inteligência artificial não é diferente. O que pode ajudar, quando estamos tentando entender a natureza de um objeto ou fenômeno — ou o que ele *é* — é responder primeiro qual o seu objetivo. Tradicionalmente, a definição do propósito da inteligência artificial tem sido vista por dois prismas dicotômicos: o modelo racional x modelo humano e o baseado em pensamento x baseado em comportamento.

No prisma “racional x humano”, buscamos a resposta se o propósito da máquina é mimetizar o ser humano ou alcançar um modelo de racionalidade ideal. Já sob o prisma do “pensamento x comportamento”, o que se busca é definir se a máquina tem como propósito o pensamento ou a ação. A interseção desses prismas resulta em quatro possíveis opções: o propósito da I.A. deve ser mimetizar o pensamento humano ou alcançar um ideal utópico de pensamento teórico racional? Ou então a pergunta deveria ser se o seu propósito é agir como os humanos agem ou agir a partir de um ideal de racionalidade? As possíveis respostas podem ser representadas como no quadro abaixo:

Quadro 1 – Prismas da finalidade da I.A.

	Modelo humano	**Modelo racional**
Baseado em pensamento	Sistemas que pensam como humanos.	Sistemas que pensam racionalmente.
Baseado em comportamento	Sistemas que agem como humanos.	Sistemas que agem racionalmente.

Fonte: Autoria própria.

O modelo humano, tanto baseado em pensamento, quanto em comportamento, parece ser impraticável. Isso é, se considerarmos que o

propósito da máquina é ser o mais parecido com o ser humano possível, a primeira pergunta que podemos fazer é: qual ser humano? Notoriamente, as pessoas pensam e agem diferente, não havendo um modelo padrão de pensamento ou comportamento humano. Poderia se rebater dizendo que há uma certa margem de liberdade para um modelo médio de homem que passaria perfeitamente por um ser humano, sem ser percebido como uma máquina. Alan Turing profetizou que algo assim ocorreria, e o experimento hipotético ficou conhecido como Teste de Turing (T.T.). A ideia, em suma, é que uma máquina estaria madura o suficiente quando, ao interagir com seres humanos, estes não pudessem identificar se estavam interagindo com um homem ou com uma máquina. Dessa forma, a máquina teria passado (ou vencido) o T.T.

A segunda pergunta que podemos fazer ao modelo humano, especialmente o comportamental (como no T.T.), é: o que se quer com isso? Quais as vantagens ou utilidades de se criar uma máquina que pense e aja de forma idêntica aos seres humanos? Historicamente, todas as tecnologias foram desenvolvidas visando facilitar a vida humana, mas como pode ser que uma réplica cibernética idêntica a nós possa ser mais útil que uma réplica que não possua as nossas “falhas”? Utilizando a automação como exemplo, se ela visa substituir o ser humano em alguns tipos de trabalhos, se tal substituição for feita por máquinas que sentem cansaço, se distraem, se perdem em pensamentos sobre o futuro etc., como isso seria desejável diante da possibilidade de ter uma máquina que não aja e pense dessa forma? Como bem observado por Russell e Norvig (2021, p. 20), o projeto do avião, que é uma forma de voo artificial, só começou a progredir quando os projetistas deixaram de tentar replicar os pássaros e começaram a observar as leis da aerodinâmica da física para aplicar em seus desenhos. Afinal, de que serviria um avião que se confunde com os pássaros, de modo que estes não soubessem identificar se estariam diante de um deles ou de uma máquina?!

Como o nosso tema central é a justificativa racional da escolha, seguiremos assumindo que o principal propósito da I.A. seja agir racionalmente, afinal, escolhas são ações. Portanto, nos coube analisar a investigação filosófica sobre a escolha racional, nos capítulos anteriores, para que neste, assim como os projetistas de aviões do passado, possamos compreender as formas de

aplicação da racionalidade à programação de algoritmos inteligentes. Ressalta-se que não se quer defender, com isso, que a ação racional seja o único propósito da I.A., mas o principal. Eventualmente — e isso será abordado adiante — ações cujas definições não sejam a mais “pura racionalidade” serão comportamentos desejáveis para o comportamento da máquina, sem que isso afete o seu propósito.

3.1 AS PREFERÊNCIAS HUMANAS

A escolha racional, como vimos, depende da comparabilidade que, por sua vez, é afetada quando não há uma medida comum para se realizar a comparação. A esse fenômeno chamamos de incomensurabilidade. Em que pese haja correntes de pensamento que afirmem que a incomensurabilidade implique na incomparabilidade, vimos que existem pensadores que sustentam a possibilidade de se comparar bens incomensuráveis.

A visão da paridade, de Ruth Chang, assume que o agente racional, diante de opções incomensuráveis, tem uma liberdade de agir diante de escolhas difíceis. Usando os mesmos termos de Chang, o agente possui uma “fonte” diversa para operar a escolha, sendo essa fonte um elemento subjetivo baseado nas razões internas do próprio agente (“criadas” por ele), diferenciando-se das razões externas (“dadas” a ele). Raz tem uma visão semelhante. Para ele, quando as opções forem racionalmente elegíveis, isso é, quando todas elas podem ser consideradas racionais, o agente pode operar a escolha por meio da vontade. Isso não implica que ele entenda que escolhas são, em si, razões para agir. A vontade pode operar a escolha, em algumas situações, contudo, é o objeto da escolha que se torna um fim que, este sim, será uma razão ou fonte de outras razões.

Ao tratar especificamente da incomensurabilidade de valores, surge um ponto crítico na teoria de Raz, que é origem dessa fonte subjetiva. Para o filósofo, há uma gama de valores que devem sua existência às práticas sociais. Sua tese da dependência social dos valores reivindica que, salvo algumas exceções, os valores dependem direta ou indiretamente de práticas sociais, seja porque sua origem se deu por uma prática social, ou porque dependem de outros valores

que, em algum momento, se originaram de práticas sociais. Isso não implica que a prática social é uma condição suficiente para que um valor seja um valor, o que aproximaria sua tese de uma espécie de relativismo, mas apenas de uma condição necessária.

O ponto chave é que enquanto parte das teorias da razão prática tentam implicar que as escolhas corretas são universais, ou seja, estando dois agentes em idêntica situação fática, a resposta correta seria a mesma para ambos, as teorias de Raz e Chang admitem uma maior margem de escolha ao agente, podendo ser o caso de que, em situações idênticas, ambos possam escolher opções diferentes de forma racional. Em outras palavras, quando a racionalidade permitisse o arbítrio da vontade do agente, não haveria uma única escolha correta.

Embora as teorias de Chang e Raz possam funcionar com o agente racional humano, não parece ser o caso de funcionarem com a inteligência artificial. Isso porque as condutas "puramente racionais" são apenas uma parte da decisão, e, frequentemente, uma parte insuficiente para operar a escolha. A adição que permite a escolha racional em casos difíceis é justamente um elemento subjetivo de valor que pode variar conforme o agente. Esse elemento subjetivo, que chamaremos de “preferências”, pode apresentar alguma dificuldade em ser compreendido pela máquina, pelos motivos que serão expostos a seguir.

3.1.1 Preferências não estritamente racionais

Conforme mencionado anteriormente, assumimos que o propósito principal dos sistemas de IA seja a ação racional, porém, muitas vezes o que pode ser entendido como uma racionalidade puramente matemática, entra em conflito com valores preciosos para os seres humanos. Em outras palavras, algumas preferências humanas são como “exceções” à racionalidade pura. Nos últimos anos, alguns pesquisadores têm correlacionado o desenvolvimento das tecnologias de IA com o aumento da segregação de grupos minoritários (BENJAMIN, 2020; EUBANKS, 2018; GARCIA et al., 2020; KEYES, 2019; STARK, 2019). Um exemplo simples é o do uso de informações puramente

estatísticas para a tomada de decisão. No Brasil, segundo levantamento oficial do Departamento Penitenciário Nacional (DEPEN, 2017), 64% da população prisional é composta por pessoas negras. Isso significa que do ponto de vista estritamente estatístico, em uma escolha aleatória entre pessoas de raças diversas, é mais provável que um negro esteja em desconformidade com a lei penal do que alguém de outra raça. As razões disso não são uma causalidade entre raça e comportamento criminoso, hipótese há muito tempo refutada pela ciência, contudo, essas afirmações ainda costumam causar um desconforto, pois envolvem valores como a justiça, liberdade e igualdade de tratamento. Contudo, para fins de definição de parâmetros decisórios, uma máquina com total desconhecimento do que fundamenta os valores humanos a serem ponderados nas escolhas, poderia tomar uma decisão que trataria de forma desigual uma pessoa negra, fundamentada unicamente em dados estatísticos, sem que essa decisão fosse considerada irracional.

Longe de ser uma matéria restrita às obras de ficção científica, organizações, acadêmicos e operadores de tecnologia demonstram uma preocupação latente com os perigos de uma inteligência superdesenvolvida que não esteja alinhada com os valores humanos (CAVE, 2019; CELLAN-JONES, 2014). A preocupação com o desenvolvimento dos sistemas de IA em consonância com os valores humanos está, por exemplo, expressa na Recomendação do Conselho de Inteligência Artificial da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que, em seus princípios para o gerenciamento responsável da confiabilidade de IA, dispõe “valores centrados no ser humano e equidade”, descrevendo da seguinte forma:

> Os atores de IA devem respeitar o império do direito, os direitos humanos e os valores democráticos, por todo o ciclo do sistema de IA. Isso inclui a liberdade, dignidade e autonomia, privacidade e proteção de dados, não-discriminação e igualdade, diversidade, equidade, justiça social e os direitos do trabalho reconhecidos internacionalmente. (OECD, 2019).

Parece haver um consenso entre os atores de IA e as partes interessadas[37], no sentido de programar os sistemas de IA em consonância com

[37] Utilizo da mesma definição da OCDE, sendo atores "aqueles que desempenham um papel ativo no ciclo de vida de um sistema de IA, incluindo organizações e indivíduos que a implantam

os valores buscados pelo homem. Nesse aspecto, retomando ao quadrante “comportamento/pensamento x humano/racional” do início do capítulo, a maximização da racionalidade da máquina precisa levar em consideração aspectos da natureza humana como a moralidade.

Contudo, antes mesmo de nos perguntarmos se os valores em questão são os mesmos para todas as pessoas afetadas, é preciso, de alguma maneira, adentrar na *definição* de tais valores. Como é possível que valores possam ser expressos em termos matemáticos para que o computador possa aprendê-los?

3.1.2 Dinamismo e individualidade das preferências

Outra dificuldade que pode ser encontrada pelos algoritmos de IA ao lidarem com preferências, diz respeito à sua individualidade e dinamismo. Conforme visto no tópico 2.2, a ausência de dificuldade em comparar elementos que possuem medidas cardinais em comum era o fato de que tais medidas são universais, objetivas e independentes da vontade do agente que as mede. Vimos adiante, que estando ausente essas características, o agente poderia tomar uma decisão racional baseado em suas preferências. Entretanto, as preferências são subjetivas, particulares e dinâmicas (mudam com o tempo). Digamos que, todas as coisas consideradas, avaliamos que cerveja e vinho estão em paridade em relação ao sabor. Dessa forma, duas pessoas diferentes podem escolher racionalmente opções diferentes, sendo que uma prefere um *pint* de uma *pale ale* e a outra um cálice de um *cabernet sauvignon*. No futuro, contudo, um entusiasta de cerveja poderia convencer aquele que preferia vinhos a degustar diversas modalidades de cervejas, de modo que, em um momento posterior, a mesma pessoa que preferia vinhos, passasse a preferir cervejas, sem qualquer prejuízo à racionalidade de sua escolha. Isso demonstra que às preferências é “permitida” uma margem de escolha que não existe na “racionalidade estrita”.

---

e a operam", e as partes interessadas [*stakeholders*] englobando "todas as organizações e indivíduos envolvidos ou afetados por sistema de IA, direta ou indiretamente".

3.2 DEFININDO VALORES

Algumas preferências humanas são meros caprichos, como a preferência que alguém possa ter por uma sobremesa de papaia com cassis à uma bola de sorvete de creme, ou como o exemplo acima da escolha entre o sabor do vinho e da cerveja. Entretanto, outras preferências possuem alto grau de relevância para o agente, como o direito de se expressar livremente sem ser censurado, ou como o exemplo acima sobre a igualdade de tratamento. Essas preferências mais relevantes chamaremos simplesmente de “valores”.

O primeiro problema que lidaremos, ao tratarmos de valores “descobertos” por algoritmos de inteligência artificial, diz respeito a sua definição formal para fins de decisões. Utilizando o exemplo de 3.1.1, digamos que uma máquina — um robô policial — é desenvolvida para maximizar a segurança de um determinado ambiente e que, para atingir tal finalidade, tenha a permissão de agir ostensivamente e antecipadamente contra as pessoas em tal ambiente, de acordo com a probabilidade de dano apresentada na situação concreta. Sendo assim, diante da aferição de um comportamento potencialmente lesivo aos demais, como a identificação de alguém sacando uma pistola da cintura, a máquina poderia disparar uma substância neutralizadora no criminoso em potencial. Imaginando uma situação menos dramática, em que ninguém apresente comportamento potencialmente lesivo, tendo em vista que a máquina fora desenvolvida para maximizar a segurança, evitando que eventos danosos ocorram, há ainda situações em que poderia racionalmente cometer uma injustiça ou, em outras palavras, violar um valor humano em favor da segurança. Se a máquina foi abastecida com dados estatísticos de crimes, poderia concluir que a presença de determinas etnias acarretaria um risco estatístico de lesão aos demais. O intrigante é que tal decisão não é baseada em uma falta de informação, pois ela estaria estatisticamente correta[38], contudo, a remoção de certos grupos sociais de um local, fundamentado somente em dados estatísticos,

[38] Uma possível hipótese que pode explicar a proporção entre negros e brancos em desconformidade com a lei é a de que ainda existiria uma consequência socioeconômica deletéria oriunda do período da escravidão, o que explicaria a proporção entre negros e brancos em situação de miséria. Sendo assim, a causalidade viria da correlação entre miserabilidade e crime, e não entre raça e crime. Porém, tal informação é irrelevante para fins de segurança, no exemplo mencionado, pois estatisticamente, a remoção de grupos minoritários do ambiente efetivamente diminuiria a probabilidade de crimes.

implicaria em uma violação a um valor que é reconhecidamente precioso, qual seja, a igualdade de tratamento.

No caso hipotético, o desenvolvedor poderia, ciente desse potencial problema, adicionar uma exceção ao código da máquina, para que não tomasse essa atitude em especial. Isaac Asimov (2018), ao definir as três leis da robótica[39], talvez tenha sido um dos primeiros autores a defender que a inteligência artificial deveria dar mais atenção às exceções do comportamento, do que a compreensão e codificação do comportamento em si. Nesse sentido, os estudos aa cibernética — a disciplina que estuda o comportamento de sistemas em resposta aos *feedbacks* — buscam compreender as relações entre objetivos e ações ao invés de procurar uma fórmula única que defina os valores (WIENER, 1988; VON FOERSTER, 2007). Ocorre que, no caso em análise, a adição da exceção resolveria apenas o problema previsto. Portanto, novas alterações no código, para incluir outras exceções, estariam limitadas à previsão humana de situações potencialmente conflitantes com valores que desejamos ser reconhecidos pelas máquinas a nosso serviço.

Uma linha de defesa é afirmar que esse limite de previsão humano não se aplicaria às máquinas, pois elas poderiam aprender, utilizando-se do *deep learning*, quais são os valores humanos por meio da observação profunda do comportamento das pessoas. Essa hipótese supõe necessariamente que as preferências humanas são o tipo de coisa decorrente da repetição de padrões baseados em alguma norma comportamental (ou prática social), assumindo, portanto, que são transitivas. Essa capacidade da máquina em adquirir conhecimento pode ser entendida tanto como 1) uma mera aceleração da própria capacidade cognitiva humana, caso em que as informações adquiridas seriam definíveis (pois humanamente cognoscíveis) ou 2) transcendendo o conhecimento humano, estando em um nível epistêmico superior à cognição humana, caso em que que as informações adquiridas seriam indefiníveis (pois o

[39] As três leis da robótica, de Asimov, são princípios da conduta do robô que levava em conta valores humanos no nível mais abstrato. São elas: 1. Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal; 2. Um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que entrem em conflito com a Primeira Lei e; 3. Um robô deve proteger sua própria existência, desde que tal proteção não entre em conflito com a Primeira ou Segunda Leis.

que entendemos como “definição” está subordinado aos limites da linguagem humana).

A segunda suposição pode surgir da intuição frequente que podemos ter de que a máquina nos conhece melhor que nós mesmos. Mas isso se deve, em muito, ao fato de que um sistema computacional tem uma capacidade de cálculo e memória muito superior ao nosso, especialmente em aspectos mensuráveis. O matemático mais genial do mundo pareceria medíocre se comparado a um computador moderno em uma disputa por tempo de resolução de cálculos vetoriais, mas, ainda assim, ambos estariam lidando com variáveis humanamente conhecidas. Um robô poderia nos informar que, como faremos uma viagem para a Suíça em dezembro, talvez seja melhor comprar um casaco agora, que estamos no verão brasileiro e os casacos estão mais baratos, por conta da baixa demanda. Embora isso seja altamente útil e, provavelmente, passasse despercebido se deixasse ao alvedrio de nossa própria organização pessoal, são fatos brutos, puramente matemáticos. É possível calcular a probabilidade da temperatura da Suíça em dezembro, checar se o usuário possui um casaco adequado em seu guarda-roupas, comparar as datas em que os casacos são mais baratos nas lojas e sugerir antecipadamente a compra da roupa. Essas facilidades são amplamente endossadas pelos entusiastas da internet das coisas (IoT), mas são apenas cálculos feitos com extrema eficiência e utilidade.

O problema disso é que parece que a forma como o algoritmo nos compreende parte de uma análise profunda de tudo que já fomos, já fizemos e que nos programamos conscientemente a fazer (como a viagem à Suíça). Assim, pode ser que as decisões tomadas por uma inteligência artificial estejam em maior conformidade com nossas vontades do que as decisões que tomaríamos espontaneamente. Porém, existem situações que o que irá mais se conformar com nossas razões são coisas que quebrem o padrão previsível ou de normalidade. Às vezes, como pondera Raz, somos impulsionados a tomar decisões inovadoras que, embora não sejam incoerentes com nossas ações passadas, possuem um elemento inovador imprevisível. Além disso, não é preciso que todas as nossas decisões sejam coerentes com as nossas ações anteriores para que sejam consideradas racionais.

Essa tese parece levar à ideia de uma superioridade da máquina sobre a humanidade, ainda que os próprios seres humanos não soubessem explicar perfeitamente o porquê de tal superioridade e, consequentemente, os motivos pelos quais as decisões devem ser tomadas pela máquina[40]. Em termos simples, seria assumir a conclusão de que a máquina é simplesmente superior epistemologicamente ao homem, mas sem saber *como*. Por fugir muito do escopo deste trabalho, essa hipótese não será abordada aqui.

Assumindo, portanto, a primeira hipótese, ou seja, a de que as preferências humanas podem ser definidas em termos humanamente cognoscíveis e, portanto, traduzidas em códigos de programação, adentramos nos problemas da descrição dos fenômenos, essencialmente, o problema da vagueza. Quando podemos ter uma definição precisa e conceitualmente fechada acerca de duas coisas e de como elas se relacionam, a comparação entre elas é uma tarefa simples. As escolhas se tornam difíceis na medida em que há uma incerteza acerca dos limites descritivos dos valores. Como visto no tópico 2.4, existem essencialmente três formas de abordar o problema da definição na investigação filosófica: o epistemicismo, o incomparabilismo e o indeterminismo semântico. Em seguida, serão mais bem conceituadas cada uma dessas visões, a fim de que possamos compreender se podem ser conhecidas por um algoritmo de inteligência artificial.

3.2.1 Epistemicismo

Para o epistemicismo, há um limite real entre um conceito e outro, mas não o conhecemos. Exemplificando pelo Paradoxo de Sorites, há um momento em que a remoção de um grão de areia (ou de uma fração de grão) do monte o desqualifica como monte, contudo, nossa limitação epistêmica não permite que

[40] A preocupação que uma confiança cega nas decisões tomadas por máquinas pudesse violar valores humanos é antiga nas obras de ficção. Para usar um exemplo mais contemporâneo, o compositor holandês Arjen Lucassen, nas letras de sua banda *Ayreon*, imaginou um cenário em que o presidente de um mundo cosmopolita delegou todas as decisões importantes a um grande *mainframe*, por imaginar que ele seria mais capaz e imparcial para lidar com problemas políticos, como a poluição do meio ambiente. Ocorre que o computador chegou à conclusão de que o maior problema da escassez de recursos do planeta era a espécie humana, e decidiu programar a destruição dos homens para salvar o ecossistema.

conheçamos este limite. Isso implica que, se o problema é a nossa limitação como seres humanos, é possível que a IA, por meio do *deep learning*, possa romper com os limites do conhecimento e alcançar informações que antes eram inalcançáveis pelo esforço puramente humano. Ou seja, a visão epistêmica pode endossar uma superioridade de aprendizagem da máquina em busca de um conhecimento metafísico.

Essa corrente prioriza a *preference learning*, ramo da *machine learning* que lida com o aprendizado do comportamento humano. Como disse Stuart Russell, ao formular os princípios para máquinas benéficas:

> 1. O único objetivo da máquina é maximizar a realização das preferências humanas.
> 2. A máquina está inicialmente incerta sobre quais são essas preferências.
> 3. A principal fonte de informação sobre as preferências humanas é o comportamento humano. (RUSSELL, 2019, p. 173)

Essa visão costuma ser vista com entusiasmo por alguns atores da indústria de IA, que enxergam na promessa de um conhecimento ainda não descoberto, uma oportunidade de empreendimento. Nesse sentido, podemos mencionar uma frase de Mark Zuckerberg (2015):

> Também estou curioso para saber se existe uma lei matemática fundamental subjacente às relações sociais humanas que governa o equilíbrio entre quem e o que todos nós nos importamos. Eu aposto que existe.

Embora empolgante, a visão é problemática, pois não sabemos se nosso padrão de comportamento é uma boa amostragem do que queremos alcançar, assim como não sabemos se, por meio da observação, é possível extrair todos os valores que queremos, como sociedade, preservar e endossar. Duas questões principais surgem: 1) É possível modelar o aprendizado da máquina para que ele filtre do comportamento humano, *apenas* os valores que desejamos cultivar? 2) Tudo aquilo que possa ser compreendido como um valor, pode ser efetivamente aprendido apenas pela observação?

Em relação à primeira pergunta, a preocupação é a de que a mera observação do comportamento nem sempre traduziria o comportamento que desejamos modelar para um conjunto cibernético ideal. Em outras palavras, a observação de uma grande amostragem de pessoas geraria uma quantidade

imensa de dados que conteria comportamentos repulsivos que seriam, aparentemente, desejáveis (pois repetidos com frequência). Para que a máquina pudesse diferenciar um padrão ruim, que deve ser evitado, de um padrão bom, que deve ser incentivado, deveria haver algum parâmetro desses padrões, o que invariavelmente leva ao problema da definição (onde traçar a linha do bom e o do mau).

Em relação à segunda pergunta, a questão envolve a resposta acerca de se o "idealmente racional" é sempre a resposta desejável. A depender do tipo de decisões que serão responsabilidades de um sistema de IA, pode ser que nem sempre a resposta ideal seja aquela mais favorável à parte interessada. Como defendido por Raz, às vezes, a vontade desempenha um papel fundamental nas decisões, sem que estas sejam reputadas como irracionais. Às vezes, o melhor a ser feito é ter acesso às opções racionalmente elegíveis, e escolher dentre elas aquela mais inclinada pela vontade ou pelo desejo. Além disso, alguns dos nossos valores são, essencialmente, normas. Os valores éticos, por exemplo, podem, em alguma medida, ser compreendidos como conjuntos normativos para se alcançar aquilo que é bom. A construção dessas normas, objetivando o bem, é o tipo de coisa que pode ser compreendida pela mera observação, ainda que profunda, da máquina, ou necessita de promulgação (ou *input*) humana?

Em suma, é razoável pensar que há aquilo que é matematicamente aferível e aquilo que não é, mas faz parte do que nós valorizamos como indivíduos, como no caso da igualdade, exemplificada no tópico anterior. Pode parecer racional, por uma questão de segurança, impedir que negros adentrem um recinto ou que existam políticas mais rígidas para a sua presença, por questões meramente estatísticas, mas esse é o tipo de decisão que não queremos que a máquina endosse – é uma regra de valor.

Não obstante, as amostras observadas pela máquina podem sofrer manipulação do comportamento por conta da preocupação com a vigilância (indivíduos tendem a se comportar de forma diferente quando sabem que estão sendo observados), influenciando o resultado obtido.

3.2.2 Incomparabilismo

Para o incomparabilismo, alguns tipos de conflitos de valores são impossíveis de se resolver, por serem distintos em um nível basilar. Essa visão não sustenta que o mundo é impossível de ser descrito de forma precisa ou que não pode ser plenamente compreendido pela nossa inteligência, mas de que as mais variadas definições da realidade não podem dar conta da complexidade das coisas, pois suas características mais fundamentais não são prontamente discerníveis (DOBBE et al., 2021, p. 7).

O incomparabilismo se alicerça, em alguma medida, no pluralismo de valores: a ideia de que existe mais de um valor em um nível fundamental, não sendo possível reduzir todas as instâncias do que é valioso a um valor único. Por exemplo, uma versão do pluralismo político é baseada em um pluralismo moral, e alega que os diversos valores morais justificariam um sistema político liberal (MASON, 2018). Ele vai além da ideia de que as pessoas diferentes valorizam coisas diferentes, pois assume que os valores são variados de uma forma indeterminada e incomensurável, não existindo qualquer modelo teórico moral que consiga definir todas as relações possíveis entre os diversos tipos de valores existentes (MACASKILL, 2013).

3.2.3 Indeterminismo semântico

Para o indeterminismo semântico, os limites que buscamos se encontram na linguagem. Assim, os limites de definibilidade de um conceito serão os mesmos em que é possível descrevê-lo pela linguagem em uma determinada comunidade. É a visão mais comumente associada ao conceito de vagueza e ao problema central do Paradoxo de Sorites, em que os conceitos caem em casos limítrofes.

O padrão normativo que levasse em conta o paradigma do indeterminismo semântico seria baseado em uma lógica difusa, aquela que, diferente da lógica booleana, aceita gradações para os valores opostos, como um *continuum* de valores possíveis entre o verdadeiro e o falso. Por exemplo, o conceito de “bonito” podendo aceitar gradações verdadeiras como “razoável”, “aceitável”, “lindo” etc. Esse conceito foi introduzido pela teoria dos conjuntos nebulosos de Lotfi Zadeh

(1965), em que os elementos dos conjuntos possuem graus de pertinência. O conjunto de gradações compreende o intervalo em conjunto real [0,1], em que 0 representa “totalmente falso” e 1 representa “totalmente verdadeiro”, sendo os demais valores entre eles uma verdade parcial, ou seja, graus intermediários de verdade (CINTULA et al., 2021).

3.3 DINAMISMO E INDIVIDUALIDADE DOS VALORES

No tópico anterior, tratamos de valores comumente compartilhados e que, em alguma medida, são considerados de grande importância pela maior parte das pessoas. Todavia, ainda que seja possível conhecer e definir todos os valores humanos, resta a questão da proporcionalidade deles. Como observou Isaiah Berlin (2002, p. 216-217), alguns valores não podem ser plenamente realizados em conjunto, pois o acréscimo em um deles implica no decréscimo do outro (chamaremos de “incompatíveis”). Nessas situações, há o problema sobre a decisão do quanto de cada valor deve ser mantido em cada caso conflitante.

Essas incompatibilidades assumem uma forma mais visível nas discussões políticas. Por exemplo, a grande maioria das pessoas admitiria que liberdade e segurança são valores que devemos preservar, entretanto, em determinadas situações, um ganho em liberdade acarretaria uma perda em segurança, e vice-versa. Se alguém considera que a segurança é o valor mais importante na sua própria vida, deve considerar alguma interferência nesse valor, para que possa exprimir alguma liberdade, pois a única forma de elevar a própria segurança ao maior limite possível implicaria na supressão de diversas instâncias da liberdade, como dirigir um carro, jantar em um local desconhecido e até mesmo interagir com pessoas e animais. Em alguma medida, o mesmo poderia ser dito do velho embate entre liberdade e igualdade.

A questão, portanto, não trata de quais valores, em abstrato, são mais ou menos importantes que outros, mas sim, em cada possível situação concreta, quanto de um valor poderia ser reconhecido em detrimento de outro, o que remete ao conceito de descontinuidade, visto em 2.3.1. Sendo assim, mesmo dentro os valores majoritariamente reconhecidos como relevantes à humanidade, existe uma subjetividade (ou individualidade) acerca das proporções da realização ou não-realização de cada valor. Em outras palavras, pessoas

distintas podem discordar sobre o quanto uma política pública, por exemplo, deveria realizar um valor e não outro.

Além disso, o tempo e outras variáveis podem influenciar a percepção de valor nas pessoas. Assim como a pessoa, no exemplo acima, passou a preferir cervejas a vinhos, alguém, enquanto jovem, pode achar que é mais importante que existam menos impostos, pois assim ele pode administrar uma maior quantidade do seu dinheiro, porém, ao envelhecer, pode começar a valorizar mais as previdências mandatórias e impostos que sustentem um estado de bem-estar social, pois isso aumenta a segurança geral, ainda que ao custo de uma menor liberdade em gerenciar o fruto de seu trabalho. A dificuldade em definir valores em termos matemáticos em um sistema, por conta de seu dinamismo, e, consequentemente, as escolhas baseadas quando há conflitos entre esses valores, também é estudada na literatura computacional em áreas como trabalho cooperativo auxiliado por computador[41] (GREENBAUM et al., 1992), *design* participativo (HALLORAN et al., 2009) e interação humano-computador (SHILTON, 2018).

Diante dessas características, não parece plausível que existam respostas ideais a serem aferidas pela IA ao decidir sobre valores, especialmente porque não está claro que esse idealismo que busca a “melhor resposta” em todas as situações seja uma condição necessária à racionalidade. Se a racionalidade permite que haja mais de uma resposta elegível, o escopo disponível ao escolhedor humano não se estende aos algoritmos, uma vez que estes não podem codificar valores.

[41] Mais usualmente referenciada pelos teóricos em sua forma original, sem tradução, como “*computer-supported cooperative work”* (CSCW).

CONCLUSÃO

As teorias que oferecem uma alternativa racional para a escolha diante de bens incomensuráveis dependem de um elemento subjetivo, que chamamos aqui simplesmente de “preferências” do escolhedor. As preferências que possuem um grau de relevância maior para o agente e que costumam ser compartilhadas pela maioria das pessoas em uma determinada sociedade, chamamos de “valor”[42]. A subjetividade e o dinamismo desses valores indicam que eles são indefiníveis, ao menos em termos humanamente cognoscíveis. Isso implica que decisões oriundas de uma fonte não-humana e programável, como é o caso das inteligências artificiais, esbarram em um problema formal, ou seja, o de estabelecer critérios precisos para a tomada de decisão quando valores estão em jogo.

O problema da definibilidade poderia ser solucionado se assumíssemos que a tecnologia da informação evoluiu de forma tão significativa que ultrapassou os limites do conhecimento humano. Em outras palavras, seria afirmar que a inteligência artificial é epistemicamente superior ao homem natural, e o que ela decide, embora não possamos entender, é o melhor para o mundo. Como essa solução levaria a uma inevitável subserviência à máquina, fomentando um estado de insegurança e medo na sociedade, deve ser descartada.

Outro impasse para o problema da definição é que os valores possuem uma certa vagueza, e a definibilidade de conceitos vagos não é um assunto encerrado na filosofia, havendo correntes epistemológicas que diferem na própria concepção do que é o conhecimento, o que afeta a possibilidade de se definir qualquer coisa.

Se a teoria de Raz estiver correta, e alguns valores humanos são possíveis de se aferir pela observação às práticas sociais de determinadas sociedades, então o seu dinamismo — a alteração de valores ou instâncias de valor conforme o tempo — e sua subjetividade podem não oferecer grandes problemas para a descoberta, ou definição, desses valores para a capacidade superior de aprendizagem do *deep learning*, entretanto, permaneceria o

[42] Para evitar criação de novas nomenclaturas, optei por chamar genericamente de valor as preferências relevantes.

problema acerca da comparabilidade. Ou seja, uma coisa é saber quais são os valores humanos, e outra é saber qual opção escolher, justificadamente, diante de valores conflitantes.

Quanto à definibilidade em termos matemáticos, talvez essa precisão seja impossível, porém, também desnecessária. Como aceitamos uma margem de imprecisão nas nossas decisões, sem que as consideremos irracionais, talvez possamos aceitar que o perfeccionismo descritivo e uma resposta certa para todos os casos não exista. A dúvida e a discordância, presentes nas decisões dependentes de valores conflitantes, podem ser parte indissociável na natureza humana, necessária tanto para o progresso tecnológico, como para a evolução dos próprios valores. Ainda que tenhamos a tendência de enxergar o robô ideal como um ser perfeito, é razoável imaginarmos que as “imperfeições” humanas, ainda que não as enxerguemos assim, sejam também valores a serem conservados por nós e pelas máquinas para que não percamos algo fundamental da nossa humanidade.

REFERÊNCIAS

ANDERSON, Elizabeth. **Value in Ethics and Economics**. Cambridge: Harvard University Press, 1995.

ANDERSON, Elizabeth. Practical Reason and Incommensurable Goods. *In*: CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 90–109, 1997.

ARISTÓTELES. **Nicomachean Ethics**. 2. ed. Indianapolis: Hackett, 1999.

ASIMOV, Isaac. **I, Robot**. London: HarperCollins Publishers, 2018.

BBC. Microsoft's Bill Gates insists AI is a threat. **BBC News**, 29 de janeiro de 2015. Disponível em: https://www.bbc.com/news/31047780. Acesso em: 17 jul. 2022.

BENJAMIN, Ruha. **Race after technology:** abolitionist tools for the New Jim Code. Polity Books, 2019. Disponível em: https://academic.oup.com/sf/article-abstract/98/4/1/5681679?redirectedFrom=fulltext. Acesso em: 21 jun. 2022.

BENTHAM, Jeremy. **An Introduction to the Principles of Morals and Legislation**. Kitchener: Batoche Books, 2000.

BERLIN, Isaiah. **Liberty Incorporating 'Four Essays on Liberty'**. New York: University of Oxford Press, 2002.

CHANG, Ruth. Introduction*. In:* CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 1–34, 1997.

CHANG, Ruth. **Comparison and the justification of choice**. University of Pennsylvania Law Review. v. 146, n. 5, p. 1569–1598, 1998.

CHANG, Ruth. The Possibility of Parity. **Ethics**, v. 112, p. 659–688, 2002.

CHANG, Ruth. Commitments, Reasons, and the Will. *In*: SHAFER-LANDAU, R. (ed.). **Oxford Studies in Metaethics**, Oxford, University of Oxford Press, v. 8, p. 74–113, 2013a.

CHANG, Ruth. Incommensurability (and incomparability)*. In:* LAFOLLETTE, H. (ed.). **The international encyclopedia of ethics**. Malden, Blackwell Publishing, 2013b.

CHANG, Ruth. **Making Comparisons Count**. New York: Routledge, 2015.

CHANG, Ruth. Hard Choices. *In:* HEIL, J. (ed.). **Journal of the American Philosophical Association**, Cambridge: Cambridge University Press, p. 1–21, 2017.

CHARLTON, William. **Aristotle Physics: Books I and II**. New York: Oxford University Press, 2006.

CINTULA, Peter et al. Fuzzy Logic. *In:* CINTULA, Peter; FERMULLER, Christian; NOGUERA, Carles (ed.). **The Stanford Encyclopedia of Philosophy**, 2021. Disponível em: https://plato.stanford.edu/entries/ logic-fuzzy. Acesso em: 06 mar. 2022.

COPP, David. **Morality, Normativity, and Society**. New York: Oxford University Press, 1995.

D'AGOSTINO, Fred. **Incommensurability and Commensuration: The Common Denominator**. Aldershot: Ashgate, 2003.

DANCY, Jonathan. **Ethics Without Principles**. Oxford: Oxford University Press, 2004.

DESCARTES, René. **Meditations on First Philosophy with Selections from the Objections and Replies**. New York: Oxford University Press, 2008.

DOBBE, Roel et al. Hard choices in artificial intelligence. **Artificial Intelligence**, v. 300, 103555. Amsterdam: Elsevier, 2021.

DWORKIN, Ronald. **Taking Rights Seriously***.* Cambridge: Harvard University Press, 1977.

DWORKIN, Ronald. Rights as Trumps. *In:* WALDRON, J. (ed.). **Theories of Rights***.* Oxford*:* Oxford University Press, p. 153–167, 1984.

EUBANKS, Virginia. **Automating inequality:** how high-tech tools profile, police and punish the poor. New York: St. Martin's Press, 2018.

FINNIS, John. **Natural Law and Natural Rights**. Oxford: Clarendon Press, 1980.

FRITZ, Kurt Von. The discovery of incommensurability by Hippasus of Metapontum. **Annals of Mathematics**, v. 46, n. 2, p. 242-264, abr. 1945. Disponível em: https://www.jstor.org/stable/1969021. Acesso em: 21 jul. 2022.

FUTURE OF LIFE INSTITUTE. **An open letter. Research priorities for robust and beneficial artificial intelligence**. Disponível em: https://futureoflife.org/ai-open-letter/. Acesso em: 17 jul. 2022.

GARCIA, Patricia et al. No: critical refusal as feminist data practice. **CSCW'20 Companion:** Conference Companion Publication of the 2020 on Computer Supported Cooperative Work an Social Computing, p. 199-202, 2020. Disponível em: https://dl.acm.org/doi/10.1145/3406865.3419014. Acesso em: 21 jul. 2022.

GERT, Joshua. **Normative Strength and the Balance of Reasons**. Philosophical Review, vol. 116, p. 533–62, 2007.

GREENBAUM, Joan; KYNG, Morten. **Design at Work:** Cooperative Design of Computer Systems. L. Erlbaum Associates Inc., 1992.

GRIFFIN, James. **Well-Being:** Its Meaning, Measurement and Moral Importance. Oxford: Clarendon Press, 1986.

GRIFFIN, James. Incommensurability: What's the Problem. *In:* CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 35–51, 1997.

HALLORAN, John et al. The value of values: resourcing co-design of ubiquitous computing. **CoDesign**, v. 5, n. 4, p. 245–273, 2009.

HARRIS, George. **Reason's Grief**. Cambridge: Cambridge University Press, 2006.

HARRIS, George. Is incomparability a problem for anyone? **Economics and Philosophy***,* v. 23, p 65–80, 2007.

HEATH, Thomas. **A history of greek mathematics**. Oxford: Clarendon Press, 2021.

HSIEH, Nien-hê. Is comparability a problem for anyone? **Economics and Philosophy**, v. 23, n. 1, p. 65-80, 2007. Disponível em: https://philpapers.org/rec/HSIIIA-2. Acesso em: 21 jul. 2022.

HSIEH, Nien-hê. Incommensurable Values. **The Stanford Encyclopedia of Philosophy**. 2021. Disponível em: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=value-incommensurable. Acesso em: 21 jul. 2022.

HYDE, Dominic. Sorites Paradox. **The Stanford Encyclopedia of Philosophy**. Diana Raffman, 2018. Disponível em: https://plato.stanford.edu/entries/sorites-paradox/. Acesso em: 30 mar. 2022.

INSTITUTO CULTURAL HUGO DE SÃO VÍTOR. **Trivium e Quadrivium:** A Doutrina das 7 Artes Liberais. Porto Alegre, 2020.

KEYES, Os. Counting the countless: why data science is a profound threat for queer people. **Real Life**, 2019. Disponível em: https://reallifemag.com/counting-the-countless/. Acesso em: 21 jul. 2022.

KUHN, Thomas. **The Structure of Scientific Revolutions**. Chicago: The University of Chicago Press, 1996.

LI et al. **Darwinian Evolution of Prions in Cell Culture**. Science, 2010.

LUKES, Stephen. Comparing the Incomparable: Trade-offs and Sacrifices. *In:* CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 184–195, 1997.

MACASKILL, William. The infectiousness of nihilism. **Ethics**, v. 123, n. 3, p. 508–520, 2013.

MASON, Elinor. Value Pluralism. **The Stanford Encyclopedia of Philosophy**. Elinor Mason, 2018. Disponível em: https://plato.stanford.edu/entries/value-pluralism. Acesso em: 06 mar. 2022.

MILL, John. **Utilitarianism**. Auckland: The Floating Press, 2009.

MILLGRAM, Elijah. Incommensurability and Practical Reasoning. *In*: CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 151–169, 1997.

NUSSBAUM, Martha. Plato on Commensurability and Desire. *In*: NUSSBAUM, M. (ed.). **Love's Knowledge**. New York: Oxford University Press, p. 106–124, 1990.

PEIRCE, Charles Sanders. Vague. *In:* BALDWIN, J. M. (ed.). **Dictionary of Philosophy and Psychology**. New York: MacMillan, 1902.

POLLOCK, John. **Contemporary Theories of Knowledge**. Lanham: Rowman & Littlefield Publishers, 1986.

PRINCE Family Paper (Temporada 5, ep. 13). **The Office** [Seriado]. Direção: Asaad Kelada et al. Produção: B. J. Novak et al. 1 DVD (22 min.) Estados Unidos: NBC, 2009.

RAWLS, John. **A Theory of Justice**. Cambridge, MA: Belknap Press, 1971.

RAZ, Joseph. **The Morality of Freedom***.* Oxford: Clarendon Press, 1986.

RAZ, Joseph. Incommensurability and Agency*. In:* CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 110–128, 1997.

RAZ, Joseph. **Practical Reason and Norms**. New York: Oxford University Press, 2002.

RAZ, Joseph; GRIFFIN, James. Mixing Values**. Aristotelian Society Supplementary Volume**, v. 65, n. 1, p. 83–118, 1991.

REGAN, Donald. Value, Comparability, and Choice. *In:* CHANG, R. (ed.) **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 129–150, 1997.

RICHARDSON, Henry. **Practical Reasoning about Final Ends**. Cambridge: Cambridge University Press, 1994.

RUSSELL, Stuart. **Human Compatible:** Artificial Intelligence and The Problem of Control. Londres: Penguin, 2019.

SARTRE, Jean-Paul. **L'existentialisme est un humanisme**. Paris: Les Edition Nagel, 1966.

SEN, Amartya. **Collective Choice and Social Welfare**. Amsterdam: Elsevier Science, 1995.

SEN, Amartya. Maximization and the act of choice. **Econometrica**, v. 65, n. 4, p. 745–779, 1997.

SEN, Amartya. Consequential Evaluation and Practical Reason. **The Journal of Philosophy**, v. 97, n. 9, p. 447–502, 2000.

SHAKESPEARE, William. **Hamlet**. New York: Hungry Minds, 2000.

SHILTON, Katie. Values and ethics in human-computer interaction, foundations, and trends. **Human-Computer Interaction**, v. 12, n. 2, p. 107–17, 2018.

SIMON, Herbert. A Behavioral Model of Rational Choice. **Quarterly Journal of Economics**, v. 69, p. 99–118, 1955.

SINNOTT-ARMSTRONG, Walter. **Understanding Arguments:** An Introduction to Informal Logic. Stamford: Cengage Learning, 2015

SLOTE, Michael. **Beyond Optimizing**. Cambridge: Harvard University Press, 1989.

SLOTE, Michael. Two Views of Satisficing. *In:* BYRON, M. (ed.). **Satisficing and Maximizing**: Moral Theorists on Practical Reason. Cambridge: Cambridge University Press, p. 14–29, 2004.

STARK, Luke. Facial recognition is the plutonium of AI. **The ACM Magazine for Students**, v. 25, n. 3, p. 50-55, 2019. Disponível em: https://dl.acm.org/doi/abs/10.1145/3313129. Acesso em: 21 jul. 2022.

STOCKER, Michael. **Plural and Conflicting Values**. Oxford: Clarendon Press, 1990.

STOCKER, Michael. Abstract and Concrete Value: Plurality, Conflict, and Maximization. *In*: CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 196–214, 1997.

STYRON, William. **Sophie's Choice**. New York: RosettaBooks, 2000.

SUNSTEIN, Cass. Incommensurability and Kinds of Valuation: Some Applications in Law. *In:* CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 234–254, 1997.

TAYLOR, Charles. Leading a Life. *In:* CHANG, R. (ed.). **Incommensurability, Incomparability, and Practical Reason**. Cambridge: Harvard University Press, p. 170–183, 1997.

TEMKIN, Larry. An abortion argument. *In:* CRISP, R., HOOKER, B. (ed.). **Well-being and morality:** essays in honor of James Griffin. Oxford: Clarendon Press, 2000.

VON FOERSTER, Heinz. **Understanding Understanding:** Essays on Cybernetics and Cognition. Springer Science & Business Media, 2007.

WIENER, Norbert. The Human Use of Human Beings**. Cybernetics and Society**, n. 320, Da Capo Press, 1988.

WIGGINS, David. Deliberation and Practical Reason. *In:* WIGGINS, D. (ed.). **Needs, Values, Truth**. Oxford: Blackwell; Aristotelian Society Series, v. 6. p. 215–238, 1987a.

WIGGINS, David. Weakness of Will, Commensurability, and the Objects of Deliberation and Desire. *In:* WIGGINS, D. (ed.). **Needs, Values, Truth**. Oxford: Blackwell; Aristotelian Society Series, v. 6. p. 215–238, 1987b.

ZADEH, Lotfi A. Fuzzy Sets. **Information and Control**, v. 8, n. 3, p. 338–353, 1965.

ZUCKERBERG, Mark. Zuckerberg Takes Questions in Facebook Session Marred by Tech Troubles. **The New York Times**, 30 de junho de 2015. Disponível em: https://bits.blogs.nytimes.com//2015/06/30/zuckerberg-takes-questions-in-facebook-session-marred-by-tech-troubles/. Acesso em: 25 maio 2022.