

UNIVERSIDADE FEDERAL DO PARANÁ

ANDRÉ DOS MARTYRES WAGECK

A VIABILIDADE DA APLICAÇÃO DA TEORIA DOS SISTEMAS INTENCIONAIS DE
DANIEL DENNETT

CURITIBA

2019

ANDRÉ DOS MARTYRES WAGECK

A VIABILIDADE DA POSTURA INTENCIONAL DE DANIEL DENNETT

Dissertação apresentada ao departamento de pós graduação em Filosofia, setor de Ciências Humanas da Universidade Federal do Paraná, como requisito parcial à obtenção de título de Mestre em Filosofia.

Orientador: Prof. Dr. José Eduardo Freitas Porcher

CURITIBA
2019

FICHA CATALOGRÁFICA ELABORADA PELO SISTEMA DE BIBLIOTECAS/UFPR –
BIBLIOTECA DE CIÊNCIAS HUMANAS COM OS DADOS FORNECIDOS PELO AUTOR

Fernanda Emanoéla Nogueira – CRB 9/1607

Wageck, André dos Martyres

A viabilidade da aplicação da Teoria dos sistemas intencionais de
Daniel Dennett. / André dos Martyres Wageck. – Curitiba, 2019.

Dissertação (Mestrado em Filosofia) – Setor de Ciências Humanas
da Universidade Federal do Paraná.

Orientador : Prof. Dr. José Eduardo Freitas Porcher

1. Dennett, Daniel Clement, 1942 – Crítica e interpretação. 2. Teoria
dos sistemas intencionais. 3. Senso comum (Psicologia). 4. Filosofia da
mente. I. Título.

CDD – 128.2



MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS HUMANAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO FILOSOFIA -
40001016039P7

ATA Nº250

ATA DE SESSÃO PÚBLICA DE DEFESA DE MESTRADO PARA A OBTENÇÃO DO GRAU DE MESTRE EM FILOSOFIA

No dia quinze de fevereiro de dois mil e dezenove às 14:00 horas, na sala Sala 603, Programa de Pós-Graduação em Filosofia/UFPR Ed. D. Pedro II 6º andar., foram instalados os trabalhos de arguição do mestrando **ANDRÉ DOS MARTYRES WAGECK** para a Defesa Pública de sua Dissertação intitulada **A viabilidade da aplicação da teoria dos sistemas intencionais de Daniel Dennett**. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em FILOSOFIA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: JOSÉ EDUARDO FREITAS PORCHER (FAJE), KLEBER CANDIOTTO (PUC/PR), LÉO PERUZZO JÚNIOR (PUC/PR). Dando início à sessão, a presidência passou a palavra ao discente, para que o mesmo expusesse seu trabalho aos presentes. Em seguida, a presidência passou a palavra a cada um dos Examinadores, para suas respectivas arguições. O aluno respondeu a cada um dos arguidores. A presidência retomou a palavra para suas considerações finais. A Banca Examinadora, então, reuniu-se e, após a discussão de suas avaliações, decidiu-se pela aprovação do aluno. O mestrando foi convidado a ingressar novamente na sala, bem como os demais assistentes, após o que a presidência fez a leitura do Parecer da Banca Examinadora. A aprovação no rito de defesa deverá ser homologada pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais do programa. A outorga do título de mestre está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, JOSÉ EDUARDO FREITAS PORCHER, lavrei a presente ata, que vai assinada por mim e pelos membros da Comissão Examinadora.

Curitiba, 15 de Fevereiro de 2019.

JOSÉ EDUARDO FREITAS PORCHER
Presidente da Banca Examinadora (FAJE)

KLEBER CANDIOTTO
Avaliador Externo (PUC/PR)

LÉO PERUZZO JÚNIOR
Avaliador Externo (PUC/PR)



MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS HUMANAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO FILOSOFIA -
40001016039P7

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em FILOSOFIA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **ANDRÉ DOS MARTYRES WAGECK** intitulada: **A viabilidade da aplicação da teoria dos sistemas intencionais de Daniel Dennett**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua aprovação no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 15 de Fevereiro de 2019.

JOSÉ EDUARDO FREITAS PORCHER
Presidente da Banca Examinadora (FAJE)

KLEBER CANDIOTTO
Avaliador Externo (PUC/PR)

LÉO PERUZZO JÚNIOR
Avaliador Externo (PUC/PR)

AGRADECIMENTOS

Agradeço aos meus estimados colegas e professores pelas trocas valiosas de críticas e comentários. Agradeço aos meus amigos do peito por possibilitarem que minha vida sempre tivesse a alegria e a leveza que são necessárias para manter o bem-estar e a sanidade na trajetória do mestrado e pela sua compreensão quando eu precisei me ausentar dos nossos encontros.

Agradeço ao amor da minha vida, Karina, que sempre esteve do meu lado e me apoiou em todas as decisões e nos momentos mais difíceis, me dando motivação e sentido na minha vida, hoje materializados no maior presente que eu poderia sonhar, a minha filha Nicole.

E, finalmente, um agradecimento especial ao meu orientador, José Eduardo Porcher, que demonstrou sua generosidade e boa vontade me acolhendo quando mal nos conhecíamos e se mantendo firme como meu orientador, confiando no meu trabalho mesmo quando não lecionava mais em Curitiba. Agradeço pelas orientações que possibilitaram o meu aprendizado e desenvolvimento no mestrado e agradeço por me dar uma referência acadêmica do que significa ser competente, lúcido e responsável.

RESUMO

A presente dissertação tem como tarefa a investigação da viabilidade da aplicação da teoria de Daniel C. Dennett sobre as crenças, chamada Teoria dos Sistemas Intencionais, a casos de irracionalidade. Com esta investigação, pretendo contribuir para dois projetos: para o debate sobre a natureza e as características das crenças e para a discussão a respeito da teoria filosófica de Dennett sobre os estados mentais. Para tanto, no segundo capítulo, exploro os princípios que guiam a utilização da Teoria dos Sistemas Intencionais analisando as principais críticas direcionadas a ela em virtude do seu particular critério de racionalidade para a atribuição de crenças e de outros estados mentais. Ao analisar estas críticas, identifico dois grandes grupos de argumentos: aqueles que se baseiam na dissimilaridade entre a psicologia do senso comum e a Teoria dos Sistemas Intencionais e aqueles baseados na inconsistência da aplicação do critério de racionalidade. Concluo que Dennett se defende das críticas direcionadas à sua teoria, mas também concluo que há uma sentida falta de sistematização para o uso do seu critério de racionalidade. Desenvolvo, portanto, no terceiro capítulo, um método para a aplicação da Teoria dos Sistemas Intencionais que resulta na sua divisão em duas partes, cada qual possuindo uma função particular para o uso da racionalidade na interpretação. Com a criação do primeiro nível, o nível de atribuição, contemplo a função da racionalidade para a redução da indeterminação da interpretação. Já no segundo nível, o nível de previsão, a racionalidade tem a sua função de sustentar a rede de crenças do sistema que fundamenta qualquer previsão ou explicação intencional. Aplico, subseqüentemente, o método desenvolvido a dois casos famosos e muito discutidos na literatura filosófica sobre crenças, uma falha de raciocínio cometida pela falácia da conjunção e um caso de delírio de Capgras. Os resultados obtidos no primeiro caso mostraram mais claramente como a Teoria dos Sistemas Intencionais pode dar conta de abarcar explicações que apontem para um funcionamento cognitivo sub-ótimo. Já os resultados obtidos pela análise do caso de delírio revelam os motivos da dificuldade encontrada pela Teoria dos Sistemas Intencionais para explicar ou prever o comportamento de pessoas nesta condição. Concluo, assim, que não puderam ser descobertos comportamentos que a psicologia do senso comum explica e que não podem ser explicados pela Teoria dos Sistemas Intencionais, corroborando a tese de Dennett. Estudos que pontuem mais precisamente a compreensão da psicologia do senso comum para estes casos podem, e devem, ser utilizados para que esta investigação possa ser levada adiante para casos de delírio.

Palavras-chave: Crença. Teoria dos sistemas intencionais. Daniel Dennett. Psicologia do senso comum. Racionalidade.

ABSTRACT

The present dissertation has the task of investigating the feasibility of the application of Daniel C. Dennett's theory of belief, called Intentional Systems Theory, to cases of irrationality. With this investigation, I intend to contribute to two projects: the debate on the nature and characteristics of beliefs and the discussion of Dennett's philosophical theory of mental states. In the second chapter, I explore the principles that guide the use of the Intentional Systems Theory by analyzing the main criticisms directed to it by virtue of its particular criterion of rationality for the attribution of beliefs and other mental states. In analyzing these criticisms, I identify two large groups of arguments: those based on the dissimilarity between common-sense psychology and Intentional Systems Theory; and those based on the inconsistency of the application of the criterion of rationality. I conclude that Dennett defends himself from criticism directed to his theory, but a lack of systematization for the use of his criterion of rationality is readily perceived. In the third chapter, I develop a method for the application of the Theory of Intentional Systems that results in its division into two parts, each with a particular function for the use of rationality in interpretation. With the creation of the first level, the level of attribution, I consider the function of rationality in reducing the indeterminacy of interpretation. At the second level, the level of prediction, rationality has the function of sustaining the belief network of the system. This belief network is responsible for grounding any prediction or intentional explanation. Subsequently, I apply the method developed to two famous and discussed cases in the literature: a failure of reasoning, committed by the fallacy of the conjunction, and a case of Capgras delusion. The results obtained in the first case showed more clearly how the intentional system theory could account for explanations that point to a sub-optimal cognitive functioning. The results obtained by the analysis of the delusion case reveal the reasons underlying the difficulty found by the intentional systems theory to explain or predict the behavior of people in this condition. I conclude, therefore, that behaviors that are better explained by the folk psychology rather than by the intentional systems theory could not be found, corroborating Dennett's thesis. Studies that more precisely point out the folk psychology's understanding of these cases of delusion can and should be used for this investigation of delusions to be carried forward.

Keywords: Belief. Intentional Systems Theory. Daniel Dennett. Folk psychology. Rationality.

LISTA DE SIGLAS

TSI – Teoria dos sistemas intencionais

PSC – Psicologia do senso comum

SUMÁRIO

1. INTRODUÇÃO	12
2. O DEBATE SOBRE CRENÇAS, A TEORIA DOS SISTEMAS INTENCIONAIS E A DELIMITAÇÃO DA RACIONALIDADE	14
2.1 INTRODUÇÃO	14
2.1.1 CARACTERÍSTICAS DA CRENÇA	15
2.2.2 TEORIAS DA CRENÇA	18
2.3 A TEORIA DOS SISTEMAS INTENCIONAIS	22
2.4 O CONCEITO DE RACIONALIDADE PARA ESTADOS MENTAIS	27
2.5 CONCLUSÃO	31
3.1 INTRODUÇÃO	33
3.2 AS CRÍTICAS DE STICH E BORTOLOTTI.....	34
3.2.1 A CRÍTICA DE STICH	34
3.2.2 A RELAÇÃO ENTRE A PSC E A TSI	35
3.2.3 A CRÍTICA DE BORTOLOTTI	37
3.2.4 A CRÍTICA DE BORTOLOTTI E A CRÍTICA DE STICH.....	42
3.2.5 DENNETT CONTRA O ARGUMENTO DA DISSIMILARIDADE	43
3.3 A DUPLA APLICAÇÃO DO CRITÉRIO DE RACIONALIDADE	47
3.3.1 A DEFESA DE DENNETT.....	49
3.3.2 AS DUAS NOÇÕES DE RACIONALIDADE EM BORTOLOTTI	51
3.4 A CONTRIBUIÇÃO DE DUB.....	57
3.4.1 A LINHAGEM QUINEANA	58
3.4.2 A LINHAGEM RYLEANA	59
3.4.3 A ARGUMENTAÇÃO DE DUB.....	59
3.4.4 A RESPOSTA DE DENNETT ÀS CRÍTICAS LEVANTADAS.....	62
3.5 PRINCÍPIOS NORTEADORES DO CRITÉRIO DE RACIONALIDADE NA TSI	64
3.5.1 O PRINCÍPIO DA REDUÇÃO DA INDETERMINAÇÃO.....	65
3.5.2 O PRINCÍPIO DA CONEXÃO HOLÍSTICA.....	67
3.5.3 UNIÃO DOS PRINCÍPIOS PARA FORMAR O CRITÉRIO DE RACIONALIDADE.....	71
3.6 CONCLUSÃO	76
4. ASPECTOS PRÁTICOS E TEÓRICOS DA APLICAÇÃO DA TSI.....	78
4.1 INTRODUÇÃO	78
4.2 BASES PARA A FORMULAÇÃO DE UM MÉTODO DE APLICAÇÃO	79
4.2.1 O NÍVEL DE PREVISÃO	83
4.2.2 O NÍVEL DE ATRIBUIÇÃO.....	93
4.2.3 CONSIDERAÇÕES IMPORTANTES	101
4.2.4 CONSIDERAÇÕES PARA CASOS DE IRRACIONALIDADE	104

4.2.5 A APLICAÇÃO DA TSI COMO UM QUEBRA-CABEÇA	105
4.2.6 RETOMADA DAS CRÍTICAS À TSI	107
4.3 MÉTODO DE APLICAÇÃO DA POSTURA INTENCIONAL	110
4.4 APLICAÇÃO PARA CASOS PARTICULARES	116
4.4.1 CASO PARADIGMÁTICO	117
4.4.2 FALHA DE RACIOCÍNIO	119
4.4.3 DELÍRIO DE CAPGRAS	129
4.4 CONCLUSÃO	137
5. CONCLUSÃO DA DISSERTAÇÃO	141
REFERÊNCIAS.....	144

1. Introdução

A presente dissertação tem como objetivo analisar a viabilidade da aplicação da teoria de Daniel C. Dennett sobre as crenças, chamada Teoria dos Sistemas Intencionais, a casos de irracionalidade. Pretendo contribuir ao importante debate sobre a natureza e as características das crenças, que têm um papel central na filosofia da mente, na filosofia da ação, na epistemologia e na psicologia, pois norteiam nossos comportamentos, nossas reações e nossas cognições. Do mesmo modo, pretendo contribuir com a discussão a respeito da teoria filosófica de Dennett sobre os estados mentais, no qual a validade da Teoria dos Sistemas Intencionais tem um papel central, visto que a partir da Teoria dos Sistemas Intencionais, Dennett fundamenta a sua teoria da intencionalidade e dos conteúdos mentais, que sustentam as suas formulações a respeito da consciência (Dennett, 1994b).

A ideia para esta dissertação surge da observação de que é especialmente relevante para uma teoria madura das crenças de que esta dê conta de explicar casos que estão na periferia do que normalmente é tomado como crença. O modo como entendemos casos-limites de crença (e.g., casos de dissonância cognitiva e comportamental) é central para o nosso entendimento sobre o que constitui uma crença, pois esclarece os limites explicativos da teoria analisada, sendo um modo de testar a sua precisão conceitual e expor as suas fraquezas escondidas pela falta de rigor. Além disso, Dennett reconhece abertamente que a Teoria dos Sistemas Intencionais tem uma dificuldade especial em lidar com casos de irracionalidade, pois simplesmente não há uma maneira direta de sua teoria prever ou explicar comportamentos irracionais.

Levando em consideração os dois pontos mencionados e o grande impacto que a teoria de Dennett teve na discussão contemporânea a respeito da natureza da crença, sendo uma das teorias mais influentes em tal debate, a Teoria dos Sistemas Intencionais se coloca como alvo digno de uma análise cuidadosa e detalhada a respeito do modo que ela se propõe a lidar com casos de pessoas que apresentam tamanha irracionalidade que a própria atribuição de crenças passa a ser duvidosa e complicada. A minha hipótese a ser corroborada na discussão da presente dissertação é a afirmação de Dennett que a TSI pode contribuir trazendo entidades como crenças e desejos que são meramente mantidas por reconhecimento mútuo e intuição compartilhada para um domínio claramente definido de entidades cujos princípios são relativamente formais e sistemáticos.

Para formar uma teoria da mente madura que pretenda embasar um estudo científico sobre o cérebro e a mente, Dennett se utiliza de uma escolha tática: explorar o que o senso comum tem a dizer a respeito da mente a partir do ponto de vista de terceira-pessoa (Dennett, 1987, p. 5). Reconhecendo a dificuldade em convencer seus leitores e adversários dessa posição, Dennett pretende que os resultados da sua teoria falem por si próprios. Me utilizo desse espírito para a metodologia do presente trabalho, buscando os resultados da aplicação da sua teoria de uma maneira que eles possam ser avaliados como evidências que corroborem ou enfraqueçam o seu modelo teórico.

Planejo alcançar esses objetivos a partir de dois passos: (1) analisar as principais críticas direcionadas à Teoria dos Sistemas Intencionais em virtude de sua potencial dificuldade em lidar com casos de irracionalidade; e (2) aplicar a Teoria dos Sistemas Intencionais a casos particulares de irracionalidade para analisar o seu poder explicativo, a sua validade e a sua precisão conceitual. Estes serão realizados nos capítulos dois e três, respectivamente. O primeiro capítulo trata-se de uma introdução aos conceitos mais importantes que serão investigados na dissertação. Neste, apresento o debate contemporâneo sobre crença e as suas principais características; a teoria da crença de Dennett e à Teoria dos Sistemas Intencionais; e as características essenciais da racionalidade e as suas principais noções relativas a estados mentais.

Analisar a viabilidade da Teoria dos Sistemas Intencionais para explicar os casos propostos de irracionalidade também é relevante com respeito à dificuldade de atribuição de crenças em situações-limite de racionalidade, tema muito discutido na literatura filosófica recente. Essas situações contemplam casos de transtornos psiquiátricos e psicológicos, como delírios ou transtornos obsessivos-compulsivos, que se poderiam se beneficiar do rigor teórico com respeito ao conceito de crença, possibilitando identificar o que seria a expressão de uma crença e o que seria a expressão de uma categoria psicológica distinta. Dessa maneira, há implicações diretas do resultado das discussões sobre estes assuntos para o tratamento desses transtornos, assim como sobre a administração de conflitos legais que podem surgir envolvendo tais casos.

Capítulo 1

2. O debate sobre crenças, a Teoria dos Sistemas Intencionais e a delimitação da racionalidade

2.1 Introdução

O primeiro capítulo desta dissertação tem o objetivo de introduzir o leitor aos principais temas desenvolvidos nesta dissertação: a crença, a TSI e a racionalidade. Primeiramente, apresentarei os pontos mais relevantes do debate contemporâneo sobre a crença. Exponho as principais características e classificações das crenças e também as abordagens teóricas mais relevantes sobre a natureza e as propriedades das crenças. Por segundo, introduzo os pontos mais importantes acerca da teoria de Dennett sobre os estados mentais. Apresento a sua teoria das crenças, a TSI, as suas características essenciais e o seu funcionamento. E, por último, introduzo o conceito de racionalidade e as suas características mais relevantes referentes aos estados mentais.

2.2 O debate sobre crenças

Fred Dretske (1995, p. 83) afirma que o termo ‘crença’ pode se referir a dois sentidos distintos. O primeiro sentido diz respeito ao estado mental particular que ocorre a quem acredita (um tipo de estado com algum conteúdo); o segundo sentido se refere ao próprio conteúdo proposicional – algo mais como um significado, que é não localizável em quem acredita. No primeiro caso, por exemplo, a crença de Ted de que a neve é branca não é a mesma que a crença de Tom de que a neve é branca. Elas ocorrem em diferentes mentes. No segundo sentido, elas são a mesma crença (i.e., o mesmo conteúdo proposicional), a saber, que a neve é branca.

Segundo Eric Schwitzgebel (2018), o sentido utilizado pela maioria dos filósofos da mente é aquele em que o termo ‘crença’ se refere à atitude que temos sempre que tomamos uma proposição como sendo o caso (i.e., como sendo verdadeira). Este tipo de atitude é chamado de *atitude proposicional*. Geralmente, as proposições são tomadas como o que quer que sejam que as sentenças expressam. Uma atitude proposicional é o estado mental de ter alguma atitude ou posição com respeito a uma proposição ou a um potencial estado de coisas em que tal proposição é verdadeira. Tomar como verdadeiro é uma das atitudes que podem ser tomadas

diante de uma proposição, podendo-se também duvidar, esperar, temer, desejar, pretender, intencionar, etc. Apesar dessas outras atitudes, discussões sobre atitudes proposicionais geralmente tomam a crença como o principal exemplo.

2.1.1 Características da crença

Crenças ocorrentes e crenças disposicionais

Segundo Schwitzgebel (2018), não é necessário refletir ativamente a respeito de uma proposição ou estado de coisas para ter-se uma crença a seu respeito. Do vasto número de crenças que uma pessoa possui, somente uma pequena fração está no primeiro plano da sua mente em determinado momento. Assim, uma distinção relevante para o entendimento do conceito de crença é a distinção entre crenças ocorrentes e disposicionais. Enquanto que o predicado ‘ocorrente’ se refere a qualquer coisa que esteja acontecendo no momento, o predicado ‘disposicional’ se refere a alguma coisa em particular que está propensa a ocorrer sob determinadas condições.

Algumas crenças estão no primeiro plano da consciência – coisas sobre as quais a pessoa está, no momento, realmente pensando. Suponha que Harry pense que as camisas xadrez são horríveis. Apenas raramente o pensamento ou julgamento de que elas são hediondas realmente vem à tona em sua mente. Quando isso acontece, ele possui uma crença ocorrente. O resto do tempo, Harry possui a crença apenas de forma disposicional. A crença ocorrente vem e vai, dependendo das circunstâncias que a provocam, enquanto que a crença disposicional se mantém relativamente estável. Outro exemplo seria que, mesmo que você nunca tenha pensado conscientemente a respeito da possibilidade de tartarugas usarem pijamas, parece certo dizer que, no entanto, você também acreditava que elas não usam pijamas antes que sua atenção fosse chamada para tal fato (Dretske, 1995, p. 83).

Crenças implícitas vs. crenças explícitas

Outra distinção importante é aquela entre crenças implícitas e explícitas. Segundo Schwitzgebel, há dois sentidos nos quais crenças podem ser implícitas. O primeiro sentido de crenças implícitas é invocado para lidar com a questão das crenças deriváveis de uma determinada proposição acreditada. Se acredito que o número de planetas no sistema solar é

menor do que 9, então parece natural dizer que acredito que o número de planetas é menor do que 10, que o número de planetas é menor do que 11, e assim por diante. Estas últimas crenças seriam deriváveis das primeiras. Diferencia-se, desse modo, crenças implícitas meramente deriváveis das crenças explícitas, que são diretamente acreditadas pelas pessoas.

O segundo sentido de crenças implícitas é aquele em que as crenças são atribuídas não porque elas são deriváveis de outras, mas porque elas são atribuídas não-explicitamente ou não-conscientemente. Esse sentido é comumente utilizado pela psicologia empírica para questões referentes a tipos de memória e conhecimento. Nesse sentido, a memória explícita envolve a lembrança consciente de alguma informação, enquanto que a memória implícita envolve a utilização desta informação sem a percepção consciente. Conhecimento que é implícito nesse sentido não será, normalmente, implícito no primeiro sentido.

Crenças *de re* e *de dicto*

Segundo Dretske (1995), crenças envolvem a implementação de conceitos: não se pode acreditar que algo é uma vaca a menos que se entenda e se tenha o conceito ‘vaca’. É possível, no entanto, ter crenças sobre vacas sem saber o que é uma vaca, no sentido de serem crenças *de re* ou demonstrativas. O exemplo dado por Dretske é que posso, por exemplo, apontar para um animal e acreditar que ele está em meu campo de visão. Se aquele animal for uma vaca, eu teria a crença, no sentido *de re*, de que uma vaca está no meu campo de visão. Se, no entanto, estivermos usando o sentido de um referencial opaco, ou *de dicto*, não se pode acreditar que uma vaca (ou em qualquer outra coisa) é uma vaca sem entender o que é uma vaca, visto que este tipo de crença é essencialmente descritivo. Pelo motivo de que os conceitos podem permanecer distintos mesmo quando sua referência é a mesma, as descrições de crenças são de caráter intencional (em oposição a extensional).

Atomismo e holismo

Um debate controverso a respeito do conceito de crença é o debate sobre a natureza holista ou atomista das crenças. Em geral, afirma Schwitzgebel (2018), a visão atomista sustenta que o conteúdo da crença não depende, de maneira geral, das crenças relacionadas a ela e, conseqüentemente, que pessoas que compreendem e aceitam sinceramente a verdade de uma mesma proposição normalmente têm exatamente a mesma crença. A visão holista, em contrapartida, sustenta que o conteúdo de qualquer crença depende, em larga medida, de uma

gama de crenças relacionadas. Sob tal perspectiva, duas pessoas dificilmente teriam a mesma crença.

De acordo com Schwitzgebel, um ponto fraco da visão atomista é não levar em conta a influência que as crenças têm umas sobre as outras, algo que parece natural. Se Ani e Sanjay acreditam na proposição “o salmão é um peixe”, mas apenas um deles sabe que o salmão nada contra a corrente e é predado por ursos no Alasca, não diríamos que eles têm exatamente a mesma crença. A visão holista, por outro lado, peca ao sugerir uma divisão arbitrária entre o que caracterizaria crenças iguais, à medida que elas divergem pela sua relação com outras crenças. A conclusão que se chega, afirma Schwitzgebel, é que a similaridade entre crenças é uma questão de grau.

Grau de confiança

O fato de que crenças aparecem em graus não somente se refere à similaridade entre crenças, mas também se refere à sua confiança. Na verdade, poderíamos supor que um sujeito sustenta cada uma de suas crenças com um nível particular de convicção. Quanto maior é a confiança em uma proposição, maior é a sua inclinação para usá-la como embasamento para suas ações. Uma forma comum de formalizar esta ideia é utilizar uma escala que varia dentro de um contínuo de 0 até 1, em que 0 indica a certeza na falsidade de uma proposição, 1 indica a certeza em sua verdade e 0,5 indica a mesma confiança de que a proposição seja verdadeira ou seja falsa (i.e., a suspensão do juízo).

Tal estratégia probabilística pode ser melhor entendida pelo reflexo que a confiança de uma pessoa em determinadas proposições ou estados de coisa teria em seus comportamentos no contexto de uma aposta. O grau de confiança em uma crença seria, assim, equivalente ao máximo que uma pessoa estaria disposta a apostar em uma aposta que não paga nada se a proposição for falsa e paga uma unidade se a proposição for verdadeira. Assim, por exemplo, se o sujeito pensa que a proposição “o restaurante está aberto” é três vezes mais provável de ser verdadeira do que falsa, ele deveria estar disposta a pagar não mais do que R\$ 0,75 por uma aposta que não paga nada se o restaurante estiver fechado e paga R\$ 1,00 se estiver aberto. Consequentemente, o grau de crença do sujeito seria 0,75 ou 75%. Tal abordagem formalizada do grau de crença provou ser útil na teoria da decisão, na teoria dos jogos e na economia.

Crença e aceitação

Ocasionalmente, outra distinção feita por filósofos, afirma Schwitzgebel, é a distinção entre crença e aceitação. Em contraste com a crença, de modo geral, a aceitação seria mantida sob o controle voluntário do sujeito, pois envolve a decisão de suspender a investigação e agir como se a questão estivesse encerrada. Por exemplo, um cientista quando confrontado com evidências que apoiam uma teoria, evidências reconhecidamente não decisivas para comprová-la, pode escolher aceitar a teoria ou não aceitá-la. Se a teoria é aceita, o cientista deixa de investigar a sua verdade e se torna disposto a fundamentar suas próprias pesquisas com base nesta teoria.

A distinção entre aceitação e crença pode ser sustentada apontando para casos em que um sujeito aceita uma proposição sem, na realidade, acreditar nela e em casos que o sujeito acredita em uma proposição sem, no entanto, aceitá-la. Um exemplo similar ao proposto seria o seguinte: na produção científica, o cientista muitas vezes não acredita na verdade literal de alguma parte da teoria da qual o seu trabalho depende, mas aceita-a como uma base adequada para a pesquisa. Um outro exemplo para o segundo caso seria alguém que genuinamente acredita que uma escada é estável, mesmo antes de verificá-la. Mas, porque depende muito dela e porque é uma boa política em geral, ele não aceita que a escada seja estável até que se tenha verificado com mais cuidado.

2.2.2 Teorias da crença

Além da caracterização a respeito do que é a crença e quais são as suas características, é importante nos situarmos dentro do debate sobre as teorias da crença. No debate contemporâneo a respeito da natureza da crença, podem ser identificados seis diferentes propostas para responder à questão. As teorias mais importantes são: representacionalismo, disposicionalismo, interpretacionismo, funcionalismo, eliminativismo e instrumentalismo. Entraremos em maiores detalhes sobre as escolas de pensamento mais relevantes aos assuntos discutidos no presente trabalho.

Representacionalismo

É comum pensar a respeito das crenças como entidades que, de alguma maneira, estão contidas na mente. De acordo com o representacionalismo, as crenças estariam contidas na mente como fatos em forma de representações, símbolos ou marcadores, que podem ser armazenados e

lembrados quando necessário. Mais precisamente, a crença seria o estado de ter estes fatos e representações armazenados na mente. É comum um representacionista considerar que uma outra característica essencial para as crenças é que elas tenham um papel causal na produção de comportamentos. Vários elementos dessa caracterização da crença já foram desafiados, mas, segundo Schwitzgebel, é provavelmente justo dizer que a maioria dos filósofos contemporâneos da mente aceita a maior parte desse quadro, que incorpora as ideias centrais da abordagem representacional.

Jerry Fodor, um dos principais proponentes da visão representacionista, sustenta que representações mentais são sentenças em uma linguagem interna da mente (Fodor, 1987), enquanto que outros argumentam que a estrutura representativa que sustenta as crenças não é linguística. Por exemplo, há filósofos argumenta que as nossas estruturas cognitivas são mais similares a um mapa do que a uma estrutura linguística (Lewis, 1994). Sistemas representacionais baseados nesse modelo são produtivos e sistemáticos. Um terceiro modo de pensar sobre as estruturas representacionais é oferecido pelo conexionismo, que sustenta que a cognição funciona pela ativação contínua de uma série de “nódulos” conectados por conexões com força regulável (van Gelder, 1990).

Disposicionalismo

Outra importante escola de pensamento sobre as crenças é o disposicionalismo, que considera que o estado interno e as estruturas de representação podem ser apenas secundárias aos fatores que efetivamente caracterizam a crença. Perspectivas tradicionais do disposicionalismo sobre a crença afirmam que, para que alguém acredite em uma proposição, este deve ter uma determinada disposição comportamental referente a P. Muito citada também é a disposição para assentir a enunciados de P, quando a oportunidade se coloca. Outras disposições relevantes podem incluir a disposição para exibir surpresa ao se deparar com evidências para a falsidade de P, a disposição de concordar com Q se for mostrado que P implica Q e a disposição de depender da verdade de P em executar os planos de alguém. Talvez todas essas disposições possam se unir sob uma única máxima, que é, em geral, estar disposto a agir como se P fosse o caso (Ryle, 1949).

Uma questão importante que é sugerida pela discussão precedente é se estar disposto a agir de determinada maneira seria necessário para a crença em P, ou se seria meramente um sinal externo dessa crença. Segundo o disposicionalismo, afirma Schwitzgebel, os comportamentos do agente podem ser previstos mesmo que os seus mecanismos internos não

sejam conhecidos. Isto acontece pois, mesmo que seja concedido que uma criatura precise ter representações típicas humanas para se comportar completamente como um ser humano, ainda pode-se pensar que é o padrão de comportamentos efetivos e potenciais que é fundamental ao conceito de crença (i.e., que representações seriam essenciais só na medida que elas fundamentam este padrão).

Há duas objeções tradicionais ao disposicionalismo. A primeira se refere à impossibilidade de se reduzir os fatos sobre as crenças de um indivíduo a fatos sobre o seu comportamento externo, sem referência a outros estados mentais como crenças, desejos, sentimentos, etc. (Chisholm, 1957). Se isso acontecesse, segue a objeção, indivíduos que possuíssem crenças idênticas teriam, também, comportamentos idênticos, o que se sabe que não é verdade, pois o comportamento se altera de acordo com o contexto de outras crenças e desejos do indivíduo. A segunda objeção apontar para a débil relação entre comportamentos e crenças em algumas circunstâncias, como em uma pessoa paralisada, em opiniões privadas ou ainda assuntos pouco relevantes.

Segundo Schwitzgebel, diante dessas duas objeções, a maioria dos disposicionalistas se dividiu em duas classes: disposicionalistas liberais e interpretacionistas. Os disposicionalistas liberais se livram da primeira objeção se desvinculando do projeto reducionista e abarcando outros estados mentais como possíveis especificadores de disposições. Eles também ampliam a gama de disposições consideradas relevantes para a posse de uma crença, de modo a incluir pelo menos algumas disposições a experimentar episódios mentais privados (Schwitzgebel, 2002). Essa desvinculação do projeto reducionista, no entanto, deixa insatisfeito os filósofos que abordam a discussão a respeito da crença com a meta de defender o fisicalismo ou o materialismo, a visão de que tudo no mundo, incluindo a mente, é integralmente física ou material.

Interpretacionismo

De acordo com Schwitzgebel, o interpretacionismo compartilha com o disposicionalismo a ênfase nos padrões de ação e reação em vez de se focar em estruturas representacionais internas, mas mantém o foco nos comportamentos observáveis, abandonado pelos disposicionalistas liberais. Eles se diferenciam do projeto disposicionalista por endossar a indeterminação das atribuições de crença: em pelo menos alguns casos, múltiplos esquemas interpretativos incompatíveis podem ser igualmente bons, e, portanto, pode não haver nenhum fato sobre qual desses esquemas é “realmente” o correto (Davidson, 1984). Do mesmo modo, pode não existir

algo que esclareça se o sujeito “realmente” acredita em P, se a crença de que P é atribuída por um esquema, mas não pelo outro.

Funcionalismo

O funcionalismo sobre estados mentais é a teoria de que o que faz de algo um estado mental de um tipo particular é constituído por suas relações causais efetivas e potenciais com estímulos sensoriais, comportamento e outros estados mentais (Armstrong, 1968). Os funcionalistas geralmente contrastam sua visão com aquelas que dizem que o que faz de algo um estado mental de um tipo particular são os fatos sobre sua estrutura interna. A estrutura interna é relevante apenas secundariamente, na medida em que fundamenta essas capacidades causais. Assim, de acordo com o funcionalista, o que faz de um estado mental uma dor não é uma configuração neural particular. Pessoas e animais com configurações neurais muito diferentes poderiam estar igualmente com dor. O que importa é que o sujeito esteja em um estado que (grosso modo) pode ser causado por dano em seu tecido e que, por sua vez, pode causar sinais de desconforto, abstinência, fuga do estímulo doloroso e (em assuntos verbais) pensamentos e enunciados como “isso dói!” (Putnam, 1975). O funcionalismo é compatível tanto com uma abordagem representacionista da crença (como em Fodor) ou uma abordagem interpretacionista (como em Dennett).

Eliminativismo e instrumentalismo

Alguns filósofos negam a existência de crenças completamente (Churchland, 1981). Segundo esta visão, geralmente conhecida como materialismo eliminativo ou eliminativismo, a concepção cotidiana das pessoas sobre a mente, sua psicologia do senso comum (*folk psychology*), é uma teoria similar às teorias populares sobre a origem do universo ou a natureza dos corpos físicos. E, assim como nossas teorias pré-científicas sobre estes assuntos mostraram-se radicalmente erradas pela cosmologia e física científicas, também essa psicologia popular, essencialmente pré-científica, será derrubada pela psicologia científica e pela neurociência, uma vez que elas tenham avançado o suficiente, tomando o rumo de teorias e construtos obsoletos como o flogisto e o éter.

De acordo com Schwitzgebel, os instrumentalistas, de modo menos extremo, consideram as atribuições de crença úteis para certos propósitos, mas sustentam que não há fatos subjacentes definitivos sobre o que as pessoas realmente acreditam. Um tipo de

instrumentalismo, o instrumentalismo duro, nega que crenças existam em qualquer sentido. Eles sustentam, desse modo, um tipo de eliminativismo. Outro tipo de instrumentalismo, que pode ser chamado de instrumentalismo suave, mantém que as crenças são reais, mas apenas num sentido menos robusto do que normalmente é pensado. Dennett (1991) articula uma visão desse tipo.

Considere-se como uma analogia: o equador é real? Bem, não no sentido de que há uma faixa vermelha atravessando o Congo; mas dizer que um país está no equador diz algo verdadeiro sobre sua posição em relação a outros países e como ele viaja enquanto a Terra gira. As crenças são reais? Talvez não no sentido de serem representações armazenadas em algum lugar da mente; mas atribuir uma crença a alguém diz algo verdadeiro sobre os padrões de comportamento e resposta dessa pessoa. Na visão de Dennett, portanto, as crenças são tão reais quanto os equadores, os centros de gravidade, ou o “brasileiro médio”.

Vemos, assim, que a visão de Dennett se apresenta como uma forma de disposicionalismo que se afastou de suas origens, abarcando o holismo e a indeterminação para se estabelecer como uma visão interpretacionista e instrumentalista leve. Não é fácil enquadrá-lo em uma posição específica e Dennett mesmo teve dificuldades em se posicionar no debate.

My attempt to endorse and exploit this distinction as a variety of selective instrumentalism was evidently a tactical error, given the confusion it has caused. I should have forsworn the term and just said something like this: My ism is whatever ism serious realists adopt with regard to centers of gravity and the like, since I think beliefs (and some other mental items drawn from folk psychology) are like that — in being abstracta rather than part of the "furniture of the physical world" and in being attributed in statements that are true only if we exempt them from a certain familiar standard of literality. (Dennett, 1987, p. 72)

Para entender o seu posicionamento mais adequadamente, sigo agora para uma apresentação das suas principais ideias relativas ao debate sobre a crença.

2.3 A Teoria dos Sistemas Intencionais

A teoria a ser analisada no presente trabalho é a teoria sobre a atribuição de estados mentais e sobre a intencionalidade de Dennett, chamada de Teoria dos Sistemas Intencionais (doravante, TSI). A base para essa teoria foi construída em textos produzidos por Dennett desde 1971, com a sua publicação de ‘Intentional Systems’, e em conjunto com outros dois textos (Dennett, 1973, 1976) constitui a introdução para a sua obra *Brainstorms* (1978a).

Apesar de esses textos fornecerem a maior parte dos fundamentos para a sua teoria da crença, Dennett publica mais um livro com uma coletânea de textos para abordar os tópicos controversos e responder aos seus críticos. Assim, a sua obra *The Intentional Stance* apresenta poucas inovações teóricas, mas um bom número de novidades em sua exposição e defesa da sua teoria. Esse livro se propõe a ser uma visão integral com relação ao conteúdo dos estados mentais humanos, sendo que a segunda parte da sua teoria sobre a mente, a *consciência*, é explorada em outros textos e sistematizada principalmente em seu livro *Consciousness Explained* (1993b) (que não será o foco do presente trabalho).

Introdução à TSI

Segundo Dennett (1987), se tentarmos explicar os comportamentos humanos pelo mesmo método que descrevemos deslizamentos, germinação ou o magnetismo estaremos destinados a fracassar. Segundo o autor, a TSI é, acima de tudo, uma análise do significado de termos mentalistas rotineiramente utilizados pelas pessoas como “crença”, “desejo”, “expectativa”, “decisão” e “intenção”. Ao se utilizarem desses termos, elas estão utilizando a psicologia do senso comum (doravante, PSC) que usamos para interpretar, prever e explicar o comportamento de outros seres humanos, animais, artefatos como computadores ou robôs, e até nós mesmos.

A consequência ao fazermos essas interpretações parece ser a de que estamos atribuindo mente e atributos mentais àquilo que interpretamos, gerando uma série de questões sobre as condições sob as quais algo pode ser dito como dotado de uma mente, crenças, desejos e outros estados mentais. De acordo com Dennett, estes questionamentos podem ser melhor respondidos analisando-se as pressuposições e métodos das nossas práticas de atribuição. Segundo a TSI, estas práticas envolvem a aplicação de uma postura particular, um modo específico de encarar as informações disponíveis, a *postura intencional*.

A postura intencional é uma estratégia que as pessoas utilizam para interpretar o comportamento de uma entidade, tratando-a como se fosse um agente racional que controla as suas ações de acordo com os seus desejos e de acordo com o que acredita ser a melhor maneira de se chegar a eles, seguindo o que é mais razoável e apropriado (Dennett, 1971, p. 90). O termo ‘sistema’ é aqui utilizado para fazer referência a esta entidade interpretada e se refere à totalidade dos fenômenos analisados que corresponde ao agente do comportamento e (pretensão) possuidor de crenças e desejos. Ele pode ser uma pessoa, um computador, um animal, ou qualquer coisa que possa ser interpretada através da postura intencional. Qualquer sistema cujo comportamento possa ser útil e amplamente interpretado pela postura intencional é, no sentido

mais completo da palavra, um *sistema intencional*. Portanto, para que um sistema seja intencional, não é necessário nada mais do que ser amplamente previsível através da estratégia intencional (1987, p. 15).

Três posturas

As características da postura intencional são compreendidas quando comparadas com duas outras posturas com tarefas similares, a *postura física* e a *postura de projeto*. Segundo Dennett (1978), a postura física é a postura empregada pelas ciências físicas, em que se utilizam todos os conhecimentos de leis da física e da constituição física das coisas para formular uma previsão. Em geral, para coisas que não estão vivas e nem são artefatos, esta é a única postura disponível. Na verdade, qualquer coisa física pode ter o seu comportamento previsto por esta postura. Seja uma pedra, um despertador ou um peixe, prevemos que eles cairão em uma linha reta em direção ao chão ao largarmos de certa altura.

Despertadores, sendo objetos projetados (ao contrário de pedras), também são passíveis de um estilo mais sofisticado de previsão: a previsão pela postura de projeto. Podemos prever que dentro de algumas horas, por exemplo, ele emitirá um ruído alto se alguns de seus botões forem apertados na ordem certa. O interessante dessa postura é que não é preciso saber nada a respeito da sua constituição física ou dos seus mecanismos internos para explicar tal regularidade. A única coisa necessária é uma pressuposição de que o despertador foi projetado e de que ele funcionará devidamente.

Tal pressuposição, contudo, traz os seus próprios riscos para a previsão. Se o objeto não tiver sido projetado como previsto ou se ele apresentar algum defeito em seu funcionamento, a previsão fracassará e não mais será possível explicar o seu comportamento por essa postura. Porém, quando o objeto com o qual estamos lidando é complexo o suficiente, o preço moderado que se paga pela pressuposição de funcionamento correto é mais do que compensado pela enorme facilidade encontrada em sua previsão.

Por exemplo, presumamos que fosse possível prever a próxima jogada de um programa que joga xadrez pela postura física, rastreando o fluxo de elétrons pelos seus transistores e os seus efeitos físicos em seus componentes pelas leis fundamentais da física. O trabalho gigantesco dessa empreitada, no entanto, pode ser aliviado se soubermos o que cada um de seus componentes deveria realizar e como eles deveriam reagir aos códigos programados em seus sistemas. Assim, um programador poderia prever com muito menos esforço as jogadas do programa de xadrez se ele soubesse como ele foi programado (i.e., projetado). Todavia, esta

informação seria inútil se houvesse um mal funcionamento ou defeito em algum de seus componentes, forçando a redução do nível de explicação para a postura física.

Uma forma ainda mais interessante e poderosa de previsão seria a utilização da postura intencional, uma derivação da postura de projeto, em que a coisa projetada (e.g., um programa de computador que joga xadrez) é tratada como um agente, com crenças, desejos e racionalidade suficiente para fazer aquilo que deveria dado a posse destas crenças e destes desejos. Assim, uma outra maneira de prever a próxima jogada do programa de xadrez seria tratá-lo como trataríamos um adversário humano, atribuindo-lhe o desejo de vencer em conjunto a certas crenças a respeito das regras do jogo e do melhor modo de posicionar as peças a cada momento do jogo.

Essa estratégia é ao mesmo tempo mais arriscada e potencialmente recompensadora que as anteriores. Vejamos a sua utilidade ao adotá-la para lidarmos com o programa de xadrez. Excluindo as jogadas ilegais e as jogadas suicidas, obtemos uma lista de movimentos reduzida quase sem esforço. Podemos ainda notar as melhores jogadas e prever que nosso adversário irá optar por alguma delas. Essa estratégia será especialmente interessante quando há uma excelência no projeto do sistema de informações. Desse modo, podemos prever as jogadas do programa de xadrez ou os comportamentos de qualquer outro sistema que seja devidamente projetado se soubermos quais são os seus propósitos e as melhores maneiras e alcançá-los.

Portanto, a postura intencional é mais útil para lidarmos com o programa de xadrez do que as outras posturas e imprescindível ao lidarmos com outras pessoas. Dennett afirma que a postura intencional é natural aos seres humanos e esse é o único motivo de não percebermos o seu uso ubíquo em nossas vidas. Arriscamos a nossa vida todas as vezes que entramos em uma rodovia com a simples pressuposição de que os carros no sentido oposto são controlados por pessoas que querem se manter vivos.

Como a TSI funciona

Dennett explica o funcionamento da TSI na seguinte passagem:

Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many — but not all — instances yield a

decision about what the agent ought to do; that is what you predict the agent will do.
(Dennett, 1987, p. 17)

Isso significa que qualquer sistema que possa ser interpretado pela postura intencional pode ser considerado como possuidor de crenças, desejos e conteúdos mentais? Dennett pede para considerarmos o humilde termostato, um dos casos mais simples de um sistema que poderia ser concebido como intencional. Mesmo para o termostato, poderíamos conceder meia dúzia de crenças e desejos úteis para entender e prever o seu comportamento. O termostato poderia crer que a sala está gelada ou quente demais, que a caldeira está ligada ou desligada ou que, se desejar que a sala esquente, ele poderia ligar a caldeira, e assim por diante.

Dennett (1981c), no entanto, restringe essa tolerante definição de sistemas intencionais desqualificando sistemas que podem ser explicados adequadamente pelas outras posturas, ou seja, quando não há ganho de poder preditivo ou explicativo com a aplicação da postura intencional. Isso faz com que um púlpito que deseja permanecer exatamente onde está não seja classificado como um sistema intencional, pois poderíamos prever que ele ficaria parado no centro do palco pela postura física.

Por que a estratégia intencional funciona?

Segundo Dennett (1987, p. 33), a estratégia intencional funciona em seres humanos porque a seleção natural projetou os seres humanos para serem racionais, para acreditar no que eles deveriam acreditar e para desejar o que eles deveriam acreditar. No entanto, devemos separar esta pergunta de outra mais complicada, a saber, como o sistema interpretado pela postura intencional funciona, ou seja, quais são os seus mecanismos internos que fazem com que esta estratégia traga resultados positivos.

De acordo com Dennett, uma ideia popular é de que a explicação de como a estratégia funciona e a explicação de como o mecanismo funciona coincidem: para cada crença previsivelmente atribuível, haverá um estado interno funcional da máquina. Dennett afirma que aqueles que pensam que é inevitável que tal teoria se mostre verdadeira estão confundindo duas afirmações empíricas diferentes: a primeira é que a descrição intencional da postura produz um padrão objetivo e real no mundo; e a segunda é que esse padrão real é produzido por outro padrão real quase isomórfico dentro do cérebro das criaturas inteligentes. Duvidar da existência do segundo padrão real não é duvidar da existência do primeiro.

O posicionamento de Dennett no debate sobre a crença

O que constituiria a crença segundo a TSI? Crença é, de acordo com Dennett, o que um sistema intencional toma como verdade, tendo em vista os seus desejos, ao desempenhar os seus comportamentos. Segundo essa visão, crença seria um constructo lógico que, juntamente ao desejo, faz parte de um cálculo racional para a predição de comportamento de certas entidades.

My thesis will be that while belief is a perfectly objective phenomenon (that apparently makes me a realist), it can be discerned only from the point of view of one who adopts a certain predictive strategy, and its existence can be confirmed only by an assessment of the success of that strategy (that apparently makes me an interpretationist). (Dennett, 1987, p. 15)

Desse modo, Dennett se posiciona como um interpretacionista e, ao mesmo tempo, como um tipo suave de realista sobre as crenças (Dennett, 1991). Ele também se afasta do representacionalismo e se aproxima do funcionalismo afirmando que não é que atribuímos (ou deveríamos atribuir) crenças e desejos somente para coisas em que encontramos representações internas, mas sim que quando descobrimos algum objeto para o qual a estratégia intencional funciona, nos esforçamos para interpretar alguns de seus estados internos ou processos como representações internas. O que faz uma característica interna de algo uma representação poderia ser somente o seu papel em regular o comportamento de um sistema intencional (1987, p. 32). Ele também afasta a TSI de uma posição disposicionalista, a saber, do behaviorismo lógico de Gilbert Ryle, afirmando que não é possível atribuir isoladamente crenças e desejos, independentemente de outras atribuições de crenças e desejos (1987, p. 58).

2.4 O conceito de racionalidade para estados mentais

Um ponto de partida para definir a racionalidade é como uma característica apresentada por agentes cognitivos quando estes adotam crenças pelos motivos apropriados. Segundo Harold Brown (1995, p. 745), há muito tempo se sustenta que a avaliação racional requer regras rigorosas para decidir se uma proposição deve ser acreditada. A lógica formal e a matemática fornecem os exemplos mais claros de tais regras. No entanto, discussões recentes propoem que mais atenção deve ser dada às limitações cognitivas humanas, enfatizando o conceito de racionalidade limitada (do inglês *bounded rationality*) e reconhecendo um escopo considerável

de desacordo sobre o que é racional (Kahneman, 2003). Essas considerações sugerem que a capacidade humana de ser racional depende de uma capacidade básica de exercer um julgamento inteligente que não pode ser completamente capturado em sistemas de regras.

Ao encontro disso, Lisa Bortolotti (2014) afirma que não há definição de irracionalidade universalmente aceita e que pode-se perceber que a irracionalidade é um alvo em movimento. É um conceito que escapa ao esclarecimento porque o tipo de comportamento a ser sancionado e as razões pelas quais ele é sancionado mudam de acordo com os aspectos da agência humana que são considerados problemáticos em cada contexto.

Assim, Bortolotti (2010) afirma ser uma tarefa especialmente desafiadora a identificação de normas de racionalidade, pois diferentes noções de racionalidade são empregadas na literatura. Muitos distinguem entre normas de racionalidade instrumental e substantiva. As primeiras concernem, por exemplo, os processos pelos quais uma atitude é formada ou pelos quais uma decisão é alcançada. As últimas concernem, por exemplo, os fins que devem ser buscados na formação de atitudes ou na tomada de decisões.

Outra distinção comum surge da questão a respeito do critério de sucesso para a agência racional, isto é, se este é epistêmico ou pragmático. O agente racional seria aquele que acerta as coisas (por exemplo, aquele com crenças verdadeiras), ou aquele para quem as coisas correm bem (por exemplo, aquele com crenças que conduzem à felicidade ou florescimento). Há considerável sobreposição entre esses dois critérios porque, em muitos contextos, a racionalidade epistêmica é uma fonte de benefícios pragmáticos, e o bem-estar promove atitudes que contribuem positivamente para a aquisição de crenças verdadeiras. Entretanto, considerações epistêmicas e pragmáticas podem conflitar, e há crenças irracionais que são pragmaticamente benéficas e crenças racionais que levam ao fracasso.

Irracionalidade

Se há uma dificuldade notável em definir o que é racionalidade, então o que podemos dizer sobre o conceito de irracionalidade? Segundo Brown (1995), a racionalidade pode ser contraposta com a *irracionalidade* e com a *não-racionalidade*. Uma árvore ou uma pedra, por não terem a capacidade para a racionalidade, são considerados como entidades não-racionais. Porém, um ser que tem a capacidade de ser racional, mas que regularmente viola os princípios da avaliação racional, é um ser irracional. Estes dois atributos também podem ser atribuídos às crenças. Quando elas são contrárias aos ditados da razão elas são consideradas irracionais. E

são designadas como não-rationais quando elas não estão sujeitas aos ditados da razão por se referirem a preferências e gostos.

Segundo Bortolotti (2014), julgamentos de irracionalidade normalmente expressam desaprovação em relação a alguns aspectos do comportamento de agentes humanos. Agentes exibindo comportamento irracional podem ser considerados tolos, ignorantes, insensatos ou mesmo loucos. Quando seu desvio com respeito às normas racionais é significativa, até mesmo sua agência é questionada. Alguns exemplos comuns de irracionalidade são: quando as crenças do sujeito não são sustentadas pela evidência disponível; quando o raciocínio do sujeito não obedece aos padrões de raciocínio guiados por princípios básicos de lógica e probabilidade; quando as decisões do sujeito são movidas tão-somente por emoções ou instintos, e não por deliberação; quando as intenções do sujeito não refletem suas crenças e desejos; quando as ações do sujeito correspondem mal às suas intenções; quando os meios apropriados para alcançar uma meta não são selecionados pelo sujeito; quando os objetivos do sujeito não são perseguidos de forma consistente.

A partir dessas considerações, Bortolotti identifica três noções gerais de racionalidade para as crenças que correspondem a três aspectos centrais e não controversos que definem as crenças e as diferenciam de outros estados mentais: racionalidade processual, racionalidade epistêmica e racionalidade agencial.

Racionalidade processual

A racionalidade processual, segundo Bortolotti (2010), diz respeito ao modo como as crenças interagem e se relacionam umas com as outras. A ideia básica é que uma crença deve ser bem integrada com as outras crenças e os outros estados intencionais do sujeito. Por exemplo, se expresso a preferência por arroz doce em relação a mousse de chocolate e a mousse de chocolate em relação a cheesecake, esperaríamos que eu preferisse arroz doce a cheesecake. Caso contrário, o meu conjunto de preferências seria inconsistente. Violações dos princípios da lógica dedutiva e indutiva, da teoria da probabilidade e da teoria da decisão são frequentemente tomadas como instâncias de irracionalidade processual.

Segundo Bortolotti, existem na literatura muitas versões da racionalidade processual. A versão idealizada é conhecida como o cenário padrão da racionalidade (do inglês *standard picture of rationality*) (Stein, 1996, p. 4). De acordo essa versão, para um sujeito ser racional é preciso que ele tenha crenças que estejam em conformidade com os melhores padrões disponíveis de raciocínio adequado. Na literatura psicológica, versões da imagem padrão foram

usadas para avaliar o desempenho humano em tarefas de raciocínio. Outra versão de racionalidade processual advém da ideia de racionalidade mínima (Cherniak, 1986), que invoca as limitações das capacidades cognitivas humanas para a avaliação da racionalidade do sujeito.

Racionalidade epistêmica

Já a racionalidade epistêmica, afirma Bortolotti (2010), diz respeito à relação entre as crenças e a evidência disponível. Ela se estabelece quando o sujeito manifesta a capacidade de formar novas crenças firmemente fundamentadas na evidência disponível e a capacidade de atualizar suas crenças quando evidência relevante se torna disponível. Para que um sujeito seja epistemicamente racional, suas crenças não devem ser apenas sensíveis à evidência, mas também devem ser sensíveis a ela. Idealmente, hipóteses não devem ser endossadas se não houver evidência convincente que as sustente; tampouco as crenças também devem ser mantidas se houver provas irrefutáveis do seu contrário. Por exemplo, o meu comportamento é considerado como epistemicamente irracional caso eu continue a acreditar que o meu horóscopo é uma confiável fonte de previsões sobre a minha vida, mesmo que eu tenha muitas evidências contestando esta posição e mesmo que eu esteja ciente de suas previsões imprecisas.

Segundo José Bermúdez (2001), a racionalidade epistêmica governa o modo pelo qual as crenças se relacionam com a evidência e também a maneira como elas devem ser alteradas à medida que ocorrem mudanças no estado da evidência. Para Richard Foley (1987), a racionalidade epistêmica se distingue de outros tipos de racionalidade simplesmente por seu objetivo distintivo: o objetivo de acreditar em proposições verdadeiras e não acreditar em proposições falsas. Algumas questões relevantes para o entendimento adequado da racionalidade epistêmica são listadas por Philip Gerrans (2001), como: o que conta como boa evidência? Até onde devemos procurar por mais evidências? Como devemos decidir qual das duas hipóteses igualmente plausíveis devemos adotar?

Irracionalidade agencial

A racionalidade agencial, afirma Bortolotti (2010), se refere ao que um sujeito faz com suas crenças. É uma noção complexa porque reúne as dimensões teórica e prática da agência intencional. Há dois aspectos da racionalidade que podem ser destacados: (1) se o sujeito está em posição de dar razões intersubjetivamente boas para o conteúdo de uma crença relatada, e (2) se ele manifesta seu endosso ao conteúdo da crença ao agir de acordo com ela nas

circunstâncias relevantes. Exemplos de irracionalidade agencial são uma opinião relatada que o sujeito não pode justificar com boas razões ou uma crença que é comportamentalmente inerte, apesar de ter potencial para orientar as ações do sujeito.

Suponha que uma pessoa dissesse: “Meu trabalho é inútil e eu quero desistir”, mas quando questionada sobre o motivo, ela não conseguisse pensar em nenhuma razão. Isso não seria necessariamente uma falha de racionalidade se ocorresse para outros tipos de crenças ou estados intencionais cujo conteúdo não deveria ser deliberado ou justificado, mas, nesse caso, a crença relatada era um julgamento valorativo. Assim, o endosso de suas crenças pela razão era esperado da pessoa, por ser um sujeito racional. Elaborando mais o exemplo, digamos que apesar de afirmar que seu trabalho é inútil, a pessoa não mostrasse nenhuma mudança de atitude em relação ao meu trabalho. Ela vai trabalhar todos os dias com um sorriso no rosto e não se demite e nem procura por outros empregos. Este comportamento tem uma fraca correspondência com o meu relatório inicial, pois o que a pessoa faz não é explicável ou compatível com o que a pessoa diz.

2.5 Conclusão

No presente capítulo, passamos em revista às principais características da crença no debate contemporâneo sobre o assunto. As principais distinções para ter em mente ao se pensar a respeito das crenças no decorrer deste trabalho são as distinções entre crenças disposicionais e ocorrentes, entre crenças explícitas e implícitas, crenças de diferentes graus de confiança e entre crenças *de re* e *de dicto*. É importante ter em mente também o debate entre holismo e atomismo sobre a natureza da crença e a distinção entre crença e aceitação.

Vimos também que há diversas abordagens que avançam as suas próprias teorias sobre a natureza da crenças: o representacionalismo, o disposicionalismo, o interpretacionismo, o funcionalismo, o eliminativismo e o instrumentalismo. Notamos que essas visões a respeito das crenças não são necessariamente excludentes entre si, sendo que Dennett se encaixa em um interpretacionismo, em um funcionalismo e em uma forma de instrumentalismo suave.

Inspecionamos também os pontos mais importantes acerca da teoria de Dennett sobre o conteúdo dos estados mentais. A sua teoria sobre a crença, que se embasa na TSI, é uma teoria a respeito dos estados mentais encontrados dentro da PSC. Vimos que a TSI se baseia em pressuposições de crenças, desejos e racionalidade do sistema interpretado para o seu sucesso interpretativo, que é a sua própria medida de atribuição intencional adequada.

Finalmente, introduzimos algumas das características essenciais da racionalidade e passamos pelas principais noções de racionalidade quando relativas aos estados mentais: a racionalidade processual, a racionalidade epistêmica e a racionalidade agencial. Para os propósitos do presente trabalho, é essencial entender os limites da racionalidade e, para tanto, apresentei definições e exemplos de situações de irracionalidade.

Capítulo 2

O requerimento de racionalidade para a atribuição de crenças

3.1 Introdução

No presente capítulo pretendo investigar algumas das principais críticas direcionadas ao critério de racionalidade para a interpretação intencional como postulada pela TSI de Dennett. Essa investigação visa ser útil para os fins de sistematizar e explicitar a maneira mais adequada para a aplicação da postura intencional para casos particulares.

Não é incomum que críticos de Dennett falem a respeito do seu critério de racionalidade como uma grande falha em sua teoria, que não tem desculpa para estar ali, pois a afirmação de que as pessoas são necessariamente racionais para que possamos entendê-las como possuidoras de crenças ou desejos parece, para muitos filósofos, não só empiricamente errada como absurda (Dub, 2015).

Primeiramente, para lidar com essa questão, analiso as críticas de Stephen Stich elaboradas em seu artigo ‘Dennett on intentional systems’ (1981). Passo, então, para as críticas de Bortolotti e seu discernimento sobre a aplicação de critérios de racionalidade para atribuição de crença para situações-limite de racionalidade. Por fim, realizo uma análise conceitual da TSI de Dennett a partir da crítica de Richard Dub.

A teoria da intencionalidade desenvolvida por Dennett começou a ser construída e expressada de forma rudimentar em *Content and Consciousness* em 1969, foi aprimorada em ‘Intentional systems’ (1971) e, desde então, foi refinada em uma série de artigos produzidos pelo autor sobre a postura intencional e os sistemas intencionais. Reimpressões de alguns de seus artigos em seu livro *Brainstorms* em 1981 tornaram acessível o contato com uma formulação mais completa de sua teoria. Porém, Dennett afirma que a sua visão estava evoluindo em resposta às suas críticas em que ele espera revisar, reformular e estender a sua visão com a publicação de *The Intentional Stance* (1987).

A proposta do livro era, portanto, expor a TSI de modo a ser a versão final de sua formulação e resposta às críticas recebidas. Nesse livro, Dennett dedica um capítulo à reimpressão do seu artigo em resposta a Stich, ‘Making sense of ourselves’ (1981) (e uma

pequena extensão deste), com o propósito de desenvolver e expressar uma maneira adequada de explicar como a sua teoria poderia lidar com casos de irracionalidade. O presente capítulo tem como objetivo lidar com o intrincado problema de como um pressuposto de racionalidade ideal, postulado pela TSI, pode dar conta de sistemas inteligentes, mas imperfeitos e claramente não completamente racionais, como seres humanos.

Os escritos de Stich sobre o assunto são de crucial importância para a presente análise. Em seu próprio texto, Dennett aponta para esse capítulo de seu livro como aquele que lidaria com as principais dificuldades de seu critério de racionalidade ideal. Assim, o leitor poderia entender apropriadamente como esses casos controversos de aparente irracionalidade podem ser abarcados por sua teoria (1987, pps. 21, 163). se espera, portanto, uma defesa importante da teoria de Dennett contra as críticas recebidas sobre esse ponto. Para entender a defesa proposta por Dennett exploro, primeiramente, a crítica de Stich.

3.2 As críticas de Stich e Bortolotti

3.2.1 A crítica de Stich

A tese geral de Stich em ‘Dennett on intentional systems’ (1981) desafia a asserção de Dennett de que a TSI deveria substituir a PSC nas práticas de investigação científica e na construção de uma teoria madura da intencionalidade e dos estados mentais.

Stich afirma que, ao se trocar a PSC pela teoria da postura intencional, nós não seríamos capazes de dizer coisas que falamos rotineiramente uns sobre os outros. Isso aconteceria porque, ao adotarmos um requerimento de racionalidade ideal, não poderíamos dar conta das menores falhas nos nossos pensamentos ou comportamentos, assim como dar o troco errado ou sucumbir à tentação de um chocolate, pois não faria sentido atribuir falhas de inferência e inconsistência a sistemas idealmente racionais. Em outras palavras, não seríamos capazes de explicar muito do que já explicamos, a partir das práticas populares, sobre o comportamento de nossos companheiros humanos. Se pressupusermos que alguém só é capaz de realizar inferências válidas e lógicas, como deveria um bom sistema idealmente racional, não há como entender o seu comportamento ao errar (por exemplo, em seu planejamento financeiro) como apenas um erro e não um ato intencional. A conclusão lógica para uma situação em que uma pessoa erra a conta, mas é idealmente racional, é que ela errou a conta de propósito. O que não só vai contra as nossas intuições do senso comum como, segundo afirma Stich, aparenta ser uma conclusão simplesmente incorreta.

Stich nota que é comum nas práticas do senso comum que se interprete adequadamente situações de falha de inferência lógica. Não necessariamente precisamos jogar a interpretação intencional pela janela quando encontramos uma instância de irracionalidade, pois pode-se entender, como muitas vezes é o caso, que a pessoa simplesmente cometeu uma falha em seu raciocínio. Não haveria nada de incoerente em uma descrição intencional a partir do senso comum de uma pessoa que calculou erroneamente as suas finanças.

Entretanto, Stich concede que existe algo de pertinente na posição de Dennett ao afirmar que casos de irracionalidade desafiam a mera possibilidade de interpretação intencional. O que haveria de pertinente, afirma ele, é que, para casos em que a irracionalidade do agente é pouco familiar para o sujeito de interpretação, certas dificuldades para interpretá-lo intencionalmente se impõem.

Stich lida com esse cenário avançando a sua própria teoria da atribuição de estados intencionais que daria conta de explicar ambas as facilidades para entender alguns casos de irracionalidade e a dificuldade para se entender outros. Para Stich, os momentos em que o sujeito de interpretação não pode mais interpretar o comportamento de um sistema como intencional são somente os momentos em que o próprio sujeito não consegue se imaginar realizando tal erro, ou seja, o sujeito, ao se engajar em uma atividade interpretativa, realiza uma espécie de simulação em que se projeta na situação do objeto de interpretação. Isso aconteceria porque a origem da impossibilidade interpretativa se localizaria não em um critério de racionalidade por si só, mas em uma dificuldade do sujeito de interpretação em conseguir medir o comportamento e estados mentais contra si mesmo, em uma espécie de simulação do contexto interpretado.

Em outras palavras, para Stich, ao se tentar entender o comportamento de outros seres humanos, uma pessoa não se pergunta, como propõe Dennett, “Quais crenças fariam com que esse comportamento fosse racional?” mas sim “Quais crenças eu teria nessa situação se eu estivesse passando por isso e tivesse esse comportamento?”. Haveria, portanto, uma importante distinção entre comportamentos irracionais que são familiares ao sujeito e comportamentos irracionais que não são familiares.

3.2.2 A relação entre a PSC e a TSI

Ao falhar em levar em consideração o mecanismo correto pelo qual atribuímos crenças e outras atitudes aos outros, conclui Stich, a teoria de Dennett desconsideraria os casos ordinários e

incontroversos em que as pessoas, lançando mão da PSC (que Stich chama de “práticas populares”), se entendem sem maiores dificuldades ou esforços ao interpretar situações *familiares* de irracionalidade. Dessa maneira haveria uma razoável assimetria e um descompasso entre a PSC e a TSI. Essa conclusão poderia trazer grandes complicações para a TSI, caso a tese fosse de que a TSI devesse se igualar à PSC. Mas qual seria, exatamente, a relação entre a PSC e a TSI como proposta por Dennett?

A resposta para essa pergunta é central para compreender a motivação de Dennett em sua empreitada e para compreender a origem das dificuldades encontradas. Para Stich, o principal projeto que Dennett tem em mente para a TSI é de “legitimar” ou “reduzir conceitualmente” as várias noções e os variados conceitos que povoam a PSC. Essa legitimação aconteceria de modo análogo à tese de Church–Turing, que mostra como o conceito intuitivo de “procedimento eficiente” na matemática pode ser identificado com a noção formal de uma função recursiva (i.e., computável por uma máquina de Turing) (Stich, 1981, p. 43). Dennett realmente busca provar a sua tese de que todo fenômeno mental referenciado na PSC seria caracterizável (ou redutível ou legitimado) intencionalmente pela TSI, a qual forneceria uma redução do mental como entendido intuitiva e ordinariamente para um campo de entidades claramente definido, cujos princípios de organização são familiares, relativamente formais e sistemáticos, e inteiramente gerais (Dennett, 1978, pp. 30-1). Além dessa redução, o projeto de Dennett visa libertar a PSC da confusão e imprecisão associadas aos seus conceitos teóricos como ‘crença’, ‘desejos’, ‘dor’, ‘imagens mentais’ e ‘experiências’. No entanto, aponta Stich, o problema com o qual Dennett está lidando é que o encaixe entre noções intuitivas da PSC e as caracterizações que Dennett fornece para elas a partir dos sistemas intencionais não é tão confortável quanto a noção intuitiva de procedimento matemático efetivo e a noção formal de computabilidade.

Ao propor a redução da PSC à TSI e ao mesmo tempo ao aprimorar a PSC se livrando da noção comum de termos como ‘crença’ por não serem “boas entidades teóricas” (Dennett, 1978, p. xx), Dennett propõe matar dois coelhos com uma só cajadada pois existem dois projetos distintos em vista. O primeiro seria a investigação antropológica para elucidar o funcionamento mental humano e suas práticas no mundo (Dennett, 1987, p. 47), e o segundo seria buscar uma versão teórica da intencionalidade humana que possa ser coerente e precisa o suficiente para ser incorporada pelo projeto científico (Dennett, 1981, p. 41).

Stich argumenta que uma aparente contradição apareceria ao Dennett tentar reduzir a PSC à TSI e ao mesmo tempo aprimorar a PSC, livrando-se das noções de senso comum de termos como ‘crença’ por não serem boas entidades teóricas. Stich questiona se Dennett pretenderia se

valer da PSC e legitimizar os seus conceitos, ou se pretenderia alterá-los até conseguir uma teoria que fosse mais coerente e útil. A resposta, conclui Stich, é que Dennett não pretende legitimar todos os conceitos e entidades teóricas da PSC, mas somente aqueles com os quais vale a pena nos preocuparmos. Quais conceitos seriam considerados como bons e, portanto, salvos deste aprimoramento teórico? Dennett (1987) deixa claro que o critério final ainda é o mesmo e deve ser aplicado aqui novamente: o aumento da capacidade preditiva. Ou seja, a TSI não seria aquilo que a PSC inalteradamente é, mas o que é em seu melhor estado, o que merece ser levado a sério e merece ser incorporado pela ciência por ter, ao mesmo tempo, precisão conceitual e poder explicativo e preditivo.

Esse esclarecimento sobre a relação entre a TSI e a PSC mostra a vulnerabilidade da teoria de Dennett a qualquer argumento que mostre que a TSI não consegue explicar algo que a PSC consegue. Por isso Dennett precisa apresentar uma resposta bem formulada para dar conta da crítica exposta por Stich. Antes de apresentar a resposta de Dennett, no entanto, é produtivo entender as críticas avançadas por Bortolotti à TSI, as quais se unem em certa medida ao exposto por Stich, mas que apresentam as suas próprias contribuições para o debate.

3.2.3 A crítica de Bortolotti

Ávida defensora da possibilidade da atribuição de crenças e intencionalidade em situações de irracionalidade, Bortolotti também desafia e critica o critério de racionalidade proposto por Dennett. Em *Delusions and Other Irrational Beliefs* (2010), Bortolotti argumenta extensivamente a respeito do tema da natureza dos delírios, se estes são ou não crenças, ou se são algum tipo diferente de crença e o que os diferenciaria das crenças comuns de pessoas não delirantes.

Bortolotti chega à conclusão de que delírios, se são diferentes de crenças comuns, não o são pelo seu status de racionalidade, pois não há nenhum ponto claro de demarcação no contínuo de racionalidade em que uma crença passa a se tornar um delírio. Há uma sobreposição, como argumenta Bortolotti, para o mesmo tipo e intensidade de irracionalidade para delírios e para crenças comuns. Dessa maneira, delírios seriam crenças, sim, porém crenças irracionais assim como tantas outras existentes em nossa vida, desde superstições para o jogo no final de semana até erros de lógica e probabilidade.

Bortolotti argumenta que não só o critério de racionalidade não é válido para separar crenças comuns de delírios, ele também não é válido para distinguir estados doxásticos e não-

doxásticos. Assim, a sua crítica se junta à crítica de Stich ao critério de racionalidade para atribuição de crenças, afirmando que não desconsideramos em nosso cotidiano um estado que poderia ser atribuído como uma crença simplesmente por ser irracional e que, por isso, o critério de racionalidade para a atribuição de crenças não pode estar correto.

Em outros textos, Bortolotti (2004, 2005, 2015) também argumenta contrariamente à visão de que não é possível realizar um ato de interpretação quando o agente age de maneira irracional, ou seja, ela é contrária à adoção do critério de racionalidade como fator limitante para a atribuição de crenças e intencionalidade. Para entender melhor essa afirmação, elaboro brevemente o que a autora quer dizer com os termos ‘interpretação’, ‘agentes’ e ‘racionalidade’.

O conceito de interpretação se refere a uma certa teoria da interpretação intencional de um sistema, que começa com a observação do comportamento deste e termina com a atribuição de estados intencionais que possibilitam ao sujeito explicar e prever outros comportamentos do sistema. Seria, então, qualquer tentativa de fazer sentido do comportamento dos outros ao lhes atribuir estados intencionais. A ‘interpretação’ se diferencia da dita Teoria da Mente e da atribuição de crenças por ser um termo mais amplo. Teoria da Mente seria um conjunto de capacidades tipicamente encontradas em seres humanos quando eles se engajam na tarefa interpretativa, como, por exemplo, na atribuição crenças e emoções a outros seres humanos (Bortolotti, 2010).

Outro conceito importante para a análise da questão é o conceito de agência humana. Para a autora, agentes são aqueles que agem de maneira deliberada, escolhendo o melhor modo de ação para si. Isso significa que eles teriam interesses ditados pelas suas crenças e seus desejos e com base nesses interesses eles escolheriam metas para si mesmos (Bortolotti, 2014).

A racionalidade de agentes humanos, segundo Bortolotti, é caracterizada por normas. Estas normas estipulam que as crenças dos agentes devem ser verdadeiras e devem formar um conjunto coerente. Elas também estipulam que as intenções para agir devem ser governadas pela correta identificação de meios que mais provavelmente avançarão os seus fins, de modo que suas ações sejam consistentes com suas intenções (Bortolotti, 2014). Em suma, quando tentamos fazer sentido do que um outro agente está fazendo, ou prever o que ele fará em seguida, o vemos como um agente intencional e vemos as suas ações e relatos verbais como motivados por seus estados intencionais.

Podemos agora entender claramente a que Bortolotti se opõe, a saber, ao *critério de restrição de racionalidade* para a atribuição de crenças. Essa é a ideia de que a interpretação para atribuição de crenças simplesmente falha quando o comportamento é irracional. Nesta visão, a agência intencional seria a capacidade de representar o mundo e de ser guiado por esta.

Para que seres humanos sejam entendidos como seres intencionais, eles precisariam necessariamente ser agentes racionais.

Duas posições não equivalentes advêm desta visão, sendo uma delas que um intérprete não poderia atribuir crenças a sistemas que apresentam comportamentos irracionais e a outra que o intérprete não poderia atribuir crenças irracionais. A diferença está no exemplo de alguém que, por ter comportamentos irracionais, não pode ser alvo de atribuição de crenças e, por outro lado, de alguém que, mesmo tendo comportamentos incompreensíveis, não pode ter crenças irracionais atribuídas a si.

Bortolotti (2014) concede, no entanto, que há pelo menos uma correlação entre falhas de racionalidade e falhas na possibilidade de atribuição de crença e que, portanto, há uma plausibilidade *prima facie* no critério de restrição de racionalidade. Intérpretes geralmente experimentam confusão e estranheza quando confrontados com comportamentos amplamente irracionais, os quais fazem a tarefa de predição de comportamentos futuros mais árdua.

Segundo Bortolotti (2010), há diversas maneiras de desenvolver uma teoria baseada nesta visão, mas o que todas têm em comum é a racionalidade como condição necessária para a atribuição de crenças. Bortolotti postula três principais versões da restrição de racionalidade que são defendidas por quem adere a essa tese. Primeiro, uma versão mais forte, que afirma que quando padrões de racionalidade não são satisfeitos, verbalizações não podem ser atribuídas como crenças. Já as outras duas são versões mais fracas, em que a atribuição de crenças fica mais difícil quando não há um pano de fundo de racionalidade, mas que ainda assim pode ser realizada. Em uma delas há um nível mínimo de racionalidade que o indivíduo precisa alcançar enquanto que na outra a atribuição de crença variaria proporcionalmente à racionalidade encontrada.

De acordo com a segunda versão mencionada, batizada por Bortolotti como o *argumento do background*, o comportamento de um indivíduo pode ser entendido intencionalmente mesmo quando falha em satisfazer os requerimentos de uma racionalidade ideal. Seria suficiente para a atribuição de crenças que o comportamento do indivíduo fosse *em sua maior parte* (ou amplamente) racional. Este argumento se contrasta com uma versão mais fraca do critério de racionalidade que Bortolotti chama de *escala móvel (sliding scale)*. Segundo a escala móvel a possibilidade de atribuição de crenças a um indivíduo varia proporcionalmente com a sua racionalidade. Portanto para essa visão se fala da existência de crenças parciais e não somente da presença ou ausência de crenças.

A primeira e mais forte versão, em contraste com as outras duas, é atribuída principalmente a Dennett, e Bortolotti a chama de *restrição de racionalidade ideal*. De acordo

com essa versão, a interpretação depende da postulação de uma racionalidade ideal e, por conta disso, quando uma pessoa relata algo ou age de modo que não se conforma às normas da racionalidade, tal relato ou comportamento não pode ser considerado como resultado de uma crença *stricto sensu* e, portanto, a explicação deste deve abandonar termos intencionais e adotar a postura física ou de projeto, como formuladas por Dennett (1978, p. 4; 1987, pp. 16-17), com respeito ao indivíduo a ser interpretado (Bortolotti, 2004, 2010, 2015).

Embora Bortolotti não teça um contra-argumento à estratégia da escala móvel, ela faz várias críticas e considerações a respeito da restrição de racionalidade ideal de Dennett e afirma que o argumento do background seria a abordagem mais interessante e viável das três. Por acreditar ser a teoria mais viável, Bortolotti passa a maior parte do seu livro *Delusions and Other Irrational Beliefs* (2010) argumentando contra ela. Embora existam críticas mais gerais ao critério de racionalidade para atribuição de crença como um todo que são pertinentes para a TSI, as suas críticas, mesmo quando pertinentes à visão do argumento do background, não possuem a mesma força contra a TSI.

A crítica de Bortolotti ao critério de restrição de racionalidade e à TSI

As principais críticas de Bortolotti à teoria de Dennett são encontradas em *Irrationality* (2014) e *Delusions and Other Irrational Beliefs* (2010), bem como no artigo ‘Can we interpret irrational behavior?’ (2004). Segundo Bortolotti, a teoria de Dennett afirma que previsões baseadas na postura intencional são *sempre* comprometidas em casos de irracionalidade, não importando o nível de irracionalidade encontrado, mas que, de acordo com o tipo de irracionalidade apresentado, tipos diferentes de interpretações podem ser empreendidos. Se a irracionalidade é local e temporária, ocorre uma instabilidade, mas que ainda pode, por fim, repousar e favorecer uma ou outra interpretação intencional. Porém, quando a irracionalidade é sistemática e profunda, nenhum tipo de interpretação intencional poderá ser feito e o intérprete deve recorrer a níveis diferentes de explicação, como o físico e o de projeto.

Bortolotti tira as suas próprias conclusões a respeito das consequências dessa teoria sobre a intencionalidade que, segundo ela, se mostram como inviáveis. Ela ilustra o seu ponto com um exemplo. Segundo Bortolotti, Dennett seria forçado a concluir que um cenário em que alguém que é daltônico e visse o sapo Kermit (que é verde) como vermelho só poderia ser adequadamente interpretado através da postura física, pois a postura intencional já estaria fora de jogo por conta da irracionalidade apresentada pela dissemelhança entre o mundo e como a

pessoa o percebe. Porém, ainda haveria muitas maneiras de entender essa pessoa através da postura intencional e atribuir-lhe crenças significativas para previsões de seu comportamento, como, por exemplo, que esse alguém acreditaria que Kermit é da mesma cor que tomates maduros (como estes lhe aparecem). Assim, Bortolotti explica por que abandonar a postura intencional é uma atitude excessivamente drástica:

Abandoning the ship of intentional explanation and prediction before it sinks may not be the best strategy. The general worry is that even an obviously false or badly integrated belief-like state can have inferential relations with other intentional states and influence action, and thus matches the ‘job description’ of beliefs in interpretation. If ascribing such a belief-like state as a belief contributes to explanation and prediction of behaviour in intentional terms, descending to the physical stance seems to be an unjustifiably drastic move. (Bortolotti, 2014, cap. 2)

Bortolotti usa essa argumentação para inferir que o critério de racionalidade ideal é excessivo, desnecessário e prejudicial. Ela chega a tal conclusão porque argumenta que se a irracionalidade humana estiver empiricamente demonstrada e ainda assim for possível empregar a interpretação para ao menos alguns comportamentos, então a ideia de uma ligação necessária entre racionalidade e interpretação é fortemente prejudicada. Posicionando-se desta maneira, Bortolotti conclui, juntamente a Stich, que dado o aparente sucesso da interpretação popular em pelo menos alguns casos de irracionalidade, há de se considerar a hipótese de que os princípios que a guiam não são constrangidos pela racionalidade ideal, mas por outros critérios.

Bortolotti reconhece que Dennett não ignora essa dificuldade em sua teoria e que ele tenta remediá-la. Segundo ela, Dennett não fica satisfeito com a ideia de legitimar falibilidade no contexto de interpretação, mas reconhece que pessoas cujos comportamentos são previsíveis pela postura intencional podem ter crenças falsas e falhas. Bortolotti afirma que Dennett (1978) tenta resolver esse conflito permitindo flexibilidade dentro da própria postura intencional, sugerindo que a pressuposição de racionalidade poderia ser adaptada para situações particulares, baseando-se na informação disponível sobre o comportamento observado. Por exemplo, se alguém está lendo um texto cujo autor tem uma visão do mundo influenciada pela física aristotélica, esse leitor não deveria pressupor que as crenças do autor incluem a teoria da inércia e da massa relativas à física newtoniana. Intérpretes deveriam partir de um nível alto de expectativas sobre a racionalidade dos outros indivíduos, mas, à medida que erros vão aparecendo, contextos históricos e fatores de desenvolvimento devem ser agregados à interpretação e intérpretes devem revisar as suas estratégias de acordo. No entanto, assim como Stich, Bortolotti identifica que essa solução cria uma grande tensão em sua teoria referente a como lidar com o seu critério de racionalidade.

O problema, sugere Bortolotti (2014), é que quando desvios de racionalidade são acomodados dentro da postura intencional, perde-se a motivação para o apelo à racionalidade como um critério necessário e geral. Se se admite que deve-se levar em consideração a situação cognitiva do indivíduo que é objeto de interpretação, por que seria preciso comprometer-se a um princípio geral governando a interpretação?

Para Bortolotti, há diversas estratégias que podem guiar a interpretação, que ela denomina *heurísticas*. Segundo ela, o critério de racionalidade não é necessário e deve ser visto somente como uma *heurística da racionalidade*. A melhor maneira de entendê-la é meramente como um *guia* para a interpretação e não como um fator limitante, restringindo a possibilidade de atribuição intencional.

3.2.4 A crítica de Bortolotti e a crítica de Stich

Apesar de realizar uma crítica ao critério de racionalidade similar à de Stich, Bortolotti rejeita a alternativa de Stich para o problema. Segundo ela, Stich utiliza o que ela chama de *heurística de similaridade*, em que o intérprete, quando empreende uma explicação e predição do comportamento dos outros pela atribuição de estados intencionais, o faz usando a si mesmo como medida. Assim, um intérprete poderia atribuir ao locutor a crença de que existe um coelho nos arbustos porque o intérprete simula em qual estado psicológico o locutor estaria se ele mesmo estivesse naquelas condições.

Essa concepção seria problemática, no entanto, pois presume que agentes humanos podem se utilizar da heurística de similaridade por compartilharem as mesmas limitações cognitivas. Porém, não seria verdade que não se pode entender intencionalmente o comportamento de agentes que têm limitações cognitivas diferentes daquelas do intérprete. Na realidade, é sempre preciso realizar suposições a respeito de que tipos de erros os locutores teriam mais probabilidade de cometer, ajustando para parâmetros como idade, nível de educação, estilo de pensamento, inteligência e a história psicológica do locutor. Assim, Bortolotti (2014) sugere que alguma teorização deve ser realizada para dar conta das diferenças individuais, pois a suposição de que a pessoa cujo comportamento quer-se prever cometeria os mesmos erros que o intérprete não o levaria muito longe em sua busca pelo entendimento e a previsão de comportamentos do interpretado.

Desse modo, Bortolotti favorece um critério para a atribuição de crenças que se sustente em heurísticas dependentes do contexto, em vez de um único tipo, seja a heurística de racionalidade ou de similaridade, que abranja todas as situações.

3.2.5 Dennett contra o argumento da dissimilaridade

A acusação levantada por Stich, Bortolotti e outros (e.g., Webb, 1994; Nichols & Stich, 2004) de que a exigência de racionalidade é exagerada e não corresponde com o modo como a PSC lida com casos de irracionalidade, caso se sustente, levanta problemas significativos para a teoria de Dennett. Como mostra Stich, a relação entre a PSC e a TSI é importante pois faz parte da própria construção da TSI, pois se a TSI propõe ser uma redução da PSC, abarcando tudo que há de útil nela, então qualquer argumento que demonstre que há algum fenômeno que a PSC consegue explicar, mas que a TSI não, aponta para uma crucial fragilidade na teoria.

Dennett se defende dessas acusações apontando para duas falhas na argumentação. Primeiro, quando a irracionalidade se espalha, não é claro que existe qualquer interpretação estável das crenças relevantes. E, segundo, sem um background de atualizações frequentes e volumosas do preenchimento de expectativas racionais, os seres humanos, mesmo pelos mais enlouquecidos, não poderiam ter crenças irracionais atribuídas a eles, para início de conversa. O comportamento humano é simplesmente ininterpretável intencionalmente exceto enquanto a serviço de uma crença, desejo ou outro, dentro de uma lógica de racionalidade (Dennett, 2009).

Toda irracionalidade gera instabilidade

Para explicar a primeira falha nas críticas direcionadas a si, Dennett (1987) se propõe a mostrar como é possível que uma pessoa não seja inteiramente racional em todos os seus comportamentos e crenças, mas, ainda assim, possa ser interpretada através da postura intencional. Para cada instância de comportamento ou crença irracionais apresentada pelo sistema, há uma ou mais instâncias de interpretação intencional que sofre um aumento de instabilidade ou apresenta incoerência em sua descrição intencional, qualquer que seja, de crenças ou de outros estados intencionais. Assim, uma pessoa pode ser irracional em determinado momento e contexto e naquela circunstância específica não será possível interpretá-la tomando a postura intencional (ou, pelo menos, haverá um grande desafio ou instabilidade para descrevê-la em termos de crenças e desejos), havendo a necessidade de descer para a postura de projeto ou mesmo para a postura física para completar a explicação. Isso, porém, não impediria que ele pudesse ser interpretado como um sistema intencional em algum outro momento, sob algum outro aspecto. Essa instância não faria da pessoa irracional no sentido de que não pode ser interpretada intencionalmente com respeito a outros

comportamentos, mas somente no sentido de que para alguma situação, em algum contexto, ela não poderia ser compreendida intencionalmente.

Para ilustrar esse ponto, Dennett aprofunda a análise sobre o exemplo utilizado por Stich (1981) em sua crítica. O exemplo seria de um garoto que erra ao dar o troco da limonada que está vendendo em sua barraca. Apesar de seu comportamento não ser idealmente racional, afirma Stich, ele pode ser facilmente compreendido pelas práticas populares: ele simplesmente cometeu um erro na conta do troco, algo especialmente esperado para uma criança. Essa eficácia na explicação contrastaria com a dificuldade da postura intencional em gerar algum resultado.

Ao analisar-se o problema mais de perto, no entanto, Dennett argumenta que as práticas populares, na verdade, não se utilizam de uma interpretação intencional para entender o erro do menino. Apesar da suposta facilidade da compreensão possuída pelas pessoas ao lidarem com comportamentos irracionais, ou com falhas em seus raciocínios (o que Dennett não concorda que seriam comportamentos que normalmente diríamos como sendo irracionais), a mesma estratégia de descer para níveis mais básicos de explicação do comportamento seria utilizada.

Para que esse ponto fique claro, Dennett propõe tentar fazer sentido do exemplo tão-somente através de termos intencionais. Ele sugere que para que possamos entender o comportamento do garoto de não reagir como esperávamos à situação de ter dado o troco errado, como se desculpar ou suar frio por tentar enganar o seu vendedor, podemos atribuir-lhe uma crença falsa de que estava dando o troco correto. No entanto, Dennett afirma, em outras passagens (e.g., 1987, p. 18), que qualquer crença falsa atribuída ao sistema deve ser devidamente explicada por uma genealogia particular. Portanto, a questão que se impõe é: o que justificaria esse movimento interpretativo de atribuir-lhe uma crença falsa e vez de tratá-lo como um sistema racional?

É possível, como Stich sugere, que o garoto tenha cometido um erro de inferência? Dennett argumenta que esse ‘erro de inferência’ é o tipo de expressão que não é exatamente uma ação possível dentro do repertório comportamental de uma pessoa, mas sim uma interpretação retroativa que utilizamos para explicar alguns fenômenos como esquecimento gradual, erros de cálculo ou descuidos.

Digamos, sugere Dennett, que em vez de devolver 13 centavos como troco para 25 centavos do comprador na compra de uma limonada que custa 12, o menino devolva somente 11 centavos. Claramente, argumenta Dennett (1987, p. 86), essa crença não seria inferida de nenhuma de suas outras crenças, como de que “a limonada vale 12 centavos”, de que “recebeu 25 centavos”, de que “ $25 - 12 = 13$ ” e que “ $11 \neq 13$ ”. De onde a sua crença viria, então? Dennett afirma que não poderia ser de nenhum processo legitimamente intencional ou racional. Não

haveria espaço em qualquer explicação intencional para falhas de raciocínio, pois não se cometem erros por *razões* (a não ser que o próprio “erro” seja proposital, o que implicaria que na realidade não se trata de um erro). Além disso, nenhuma mera sucessão no tempo, ou mesmo causação normal, é suficiente por si só para contar como uma inferência. Portanto, precisamos descer a níveis de explicação mais baixos, pois nenhuma explicação em termos de crenças e desejos fará completo sentido do seu erro.

Resumindo a argumentação de Dennett, entender o comportamento do garoto em termos intencionais é responder à pergunta do porquê ele teria agido dessa maneira. Para responder a essa pergunta precisamos atribuir-lhe ao menos uma crença falsa, o que faz com que a irracionalidade não se apresente mais em seu *comportamento*, mas em sua *crença*. Pois, tendo em vista as suas crenças de que achava que estava dando o troco correto, o seu comportamento de dar somente 11 centavos é completamente racional. A irracionalidade foi deslocada para a sua crença falsa. Mas, novamente, para fazer sentido dela intencionalmente teríamos que atribuir uma certa inferência racional. Por exemplo, é racional que ele acredite que deu o troco correto, pois acreditava que estava devolvendo 13 centavos (e não 11). A crença de que ele daria o troco correto é salva da irracionalidade e passa para a crença de que teria devolvido 13 centavos. Esta também pode ser salva se levarmos em conta a sua crença de que havia pego três moedas, e não somente uma, para dar o troco. Mas o problema é transferido para a outra crença de que ele teria dado três moedas de um centavo. Em algum ponto a nossa descrição terá que lidar com a completa falta de sentido na transição de qualquer erro e descer a níveis mais baixos, pois nenhuma explicação em termos exclusivos de crenças e desejos fará completo sentido do seu erro. Uma hipótese necessária para fazer com que a explicação funcione seria algo como argumentar que a crença do garoto aconteceu por uma distorção ou falha em sua percepção, algo explicado pelo nível de projeto ou físico.

Dennett conclui reafirmando a sua posição, mesmo diante das críticas levantadas:

This is not to say that we are always rational, but that when we are not, the cases defy description in ordinary terms of belief and desire. (Dennett, 1987, p. 87)

So I stick to my guns: even for the everyday cases of error Stich presents, the problems of belief interpretation encountered by my view really are there in the folk-psychological practice, although they often lurk behind our confabulations and excuses. (...) This is not to say that such phenomena cannot be given any coherent description. Of course they can be coherently described from either the design stance or the physical stance — a point on which Stich and I agree. So I do not discover any truths of folk theory I must regretfully forswear. (Dennett, 1987, p. 93)

Mas se é verdade que o garoto tinha uma simples crença falsa e que isso explicaria seu

comportamento subsequente, o que fazer com essa irracionalidade? Pois, para Dennett, a mera descrição de lapsos de racionalidade em termos intencionais será problemática e trará instabilidade. A única maneira de reaver a racionalidade do sistema é por meio do background de racionalidade no qual o ato irracional está envolto.

Mistakes of the sort exhibited in this example are slips in good procedures, not manifestations of an allegiance to a bad procedure or principle. The partial confirmation of our inescapable working hypothesis that the boy is fundamentally rational is his blushing acknowledgment of his error. He doesn't defend his action once it is brought to his attention, but willingly corrects his error. This is in striking contrast to the behavior of agents in the putative cases of genuine irrationality cited by Stich. In these instances, people not only persist in their "errors," but also stubbornly defend their practice — and find defenders among philosophers as well ... It is at least not obvious that there are any cases of systematically irrational behavior or thinking. The cases that have been proposed are all controversial, which is just what my view predicts: no such thing as a cut-and-dried or obvious case of "familiar irrationality." (1987, p. 87)

Isso nos leva ao segundo contra-argumento de Dennett.

O background de racionalidade

Dennett argumenta que o comportamento irracional do agente só parece ser inteligível e interpretável intencionalmente por causa do meio de racionalidade que envolve o ato sob análise. Se não houvesse um fundo de racionalidade, o comportamento irracional não geraria somente uma instabilidade, mas sim uma total incoerência. No exemplo do menino vendedor de limonada, o argumento de que seu comportamento é compreensível intencionalmente mesmo sendo irracional lograria êxito, pois o caminho que fazemos para compreender o seu ato "irracional" é justamente através da racionalidade na ampla maioria de seus outros comportamentos.

Esperaríamos do menino, portanto, outros comportamentos e crenças racionais que possibilitariam o esclarecimento do seu comportamento irracional, como, por exemplo, o que ele faria quando apontássemos a ele o seu erro. Esperaríamos, sugere Dennett, que ele exibisse surpresa, vergonha, que desse um tapa na testa, se desculpasse e desse os dois centavos que estivessem faltando. Se ele não apresentasse estes comportamentos, ou comportamentos com funções similares, um intérprete não poderia entender claramente o seu comportamento de devolver o troco errado como uma falha claramente irracional. E a maneira pela qual se prevê tais comportamentos através da atribuição das crenças que (1) ele teria dado o troco correto e que (2) nenhum garoto de sua idade deveria errar em uma conta simples assim, bem como pela

atribuição do desejo que ele teria de não trapacear.

Em segundo lugar, não é como se a crença do garoto de que ele teria dado o troco correto fosse completamente sem fundamentos, o que tornaria a sua crença ainda mais irracional e inesperada. Não iríamos supor, por exemplo, que ele tivesse acreditado nisso caso a sua mão estivesse vazia ou cheia de moedas de 25 centavos ao devolver o troco ou caso o comprador tivesse pago a limonada com um cartão de crédito ou com um dólar. A sua irracionalidade, ou seja, sua crença falsa, é limitada a erros de confusão ou falha na sua percepção do que ele teria entregue, facilitando a interpretação intencional do caso. E, mesmo assim, precisamos descer aos níveis de projeto e físico para descrever o seu erro e como essa crença falsa teria se alojado em sua rede de crenças, pois nenhuma explicação intencional dará conta de explicar a transição em qualquer erro.

Dessa maneira, Dennett disputa vigorosamente a asserção de que os seres humanos sejam amplamente irracionais:

How rational are we? Recent research in social and cognitive psychology (e.g., Tversky & Kahneman, 1974; Nisbett & Ross 1980) suggests we are only minimally rational, appallingly ready to leap to conclusions or be swayed by logically irrelevant features of situations, but this jaundiced view is an illusion engendered by the fact that these psychologists are deliberately trying to produce situations that provoke irrational responses — inducing pathology in a system by putting strain on it — and succeeding, being good psychologists. No one would hire a psychologist to prove that people will choose a paid vacation to a week in jail if offered an informed choice. At least not in the better psychology departments. A more optimistic impression of our rationality is engendered by a review of the difficulties encountered in artificial intelligence research. Even the most sophisticated AI programs stumble blindly into misinterpretations and misunderstandings that even small children reliably evade without a second thought... From this vantage point we seem marvelously rational. (Dennett, 1987, p. 52)

3.3 A dupla aplicação do critério de racionalidade

Uma segunda crítica da qual Dennett precisa dar conta para mostrar que o seu pressuposto de racionalidade na utilização da postura intencional é pertinente trata-se da argumentação de Stich referente ao modo que Dennett pretende encaixar a PSC com a TSI. Segundo Stich, Dennett se mostra ambivalente ao expressar e explicar qual exatamente é o papel do critério de racionalidade dentro da interpretação pela postura intencional. Dennett dividiria a sua argumentação para dar conta dessa tarefa em duas linhas de explicação bem distintas, pois abordariam a questão de maneiras essencialmente diferentes. A primeira forma de se entender a questão é o que Stich chama de *linha rígida*, em que Dennett minimiza ao máximo as diferenças entre a TSI e a PSC. Segundo a linha rígida haveria, em todas as nossas

interpretações e expectativas dos outros, em nossas interações ordinárias e cotidianas, um elemento da idealização racional. Esse elemento seria fundamental para que nossa comunicação seja útil e não passe à total incoerência:

Conflict arises ... when a person falls short of perfect rationality, and avows beliefs that either are strongly disconfirmed by the available empirical evidence or are self-contradictory or contradict other avowals he has made. If we lean on the myth that a man is perfectly rational, we must find his avowals less than authoritative: "You can't mean - understand - what you're saying!" if we lean on his right as a speaking intentional system to have his word accepted, we grant him an irrational set of beliefs. Neither position provides a stable resting place; for, as we saw earlier, intentional explanation and prediction cannot be accommodated either to breakdown or to less than optimal design, so there is no coherent intentional description of such an impasse (Dennett, 1978, p. 20)

A segunda forma que Dennett empregaria para entender a questão seria a *linha suave*. Segundo ela, haveria uma diferença importante entre a TSI e a PSC quanto à necessidade da idealização de racionalidade, vendo-a mais como um parâmetro a ser buscado nas interpretações e previsões intencionais e não tanto como uma obrigatoriedade a atribuição de crenças. Stich destaca a seguinte passagem em que Dennett se utilizaria da linha suave para lidar com a irracionalidade:

Of course we don't all sit in the dark in our studies like mad Leibnizians rationalistically excogitating behavioral predictions from pure, idealized concepts of our neighbors, nor do we derive all our readiness to attribute desires to a careful generation of them from the ultimate goal of survival. ... Rationalistic generation of attributions is augmented and even corrected on occasion by empirical generalizations about belief and desire that guide our attributions and are learned more or less inductively. ... I grant the existence of all this naturalistic generalization, and its role in the normal calculation of folk psychologists — i.e., all of us. ... I would insist, however, that all this empirically obtained lore is laid over a fundamental generative and normative framework that has the features I have describe. (Dennett, 1987, p. 53)

A linha suave e a linha rígida foram delineadas por Stich e abrangem bem o que Bortolotti entende que Dennett estaria fazendo com o seu critério de racionalidade. A linha rígida foi criticada acima, mas a linha suave teria os seus próprios problemas. Para ambos, Dennett também percebe a dificuldade que a exigência de se preencher um critério de racionalidade teria sobre a interpretação, e por isso teria criado uma maneira menos exigente para lidar com casos difíceis de maior irracionalidade, empregando modos conflitantes de utilizar o critério de racionalidade. Em outras palavras, a solução de que Dennett lançaria mão para dar conta da tensão externa criada pela dissimilaridade entre a TSI e a PSC resultaria em uma tensão interna dentro da sua própria teoria. Dennett se utilizaria de critérios de racionalidade diferentes para dar conta de explicar fenômenos diferentes, criando uma inconsistência teórica.

Apesar da resposta de Dennett às questões levantadas pelos seus críticos com relação à dissimilaridade da TSI com a PSC, como considero, ser contundente, é verdade que Dennett

parece expressar critérios de racionalidade de diferentes níveis de rigor para lidar com fenômenos diferentes. Dessa maneira, apesar de montar um bom argumento para reforçar a necessidade do critério de racionalidade para atribuição de crenças e intencionalidade, Dennett ainda deixa muito no ar a respeito de como aplicar o seu critério. A acusação de Stich a respeito da inconsistência da aplicação de um critério de racionalidade, portanto, tem plausibilidade suficiente para justificar uma análise mais detalhada.

3.3.1 A defesa de Dennett

Dennett (1987, p. 94) responde diretamente a essa crítica argumentando que a divisão de sua teoria em duas maneiras diferentes de se encarar o critério de racionalidade nasce somente da frustração de Stich em tentar entender a sua teoria. Dennett, no entanto, mostra aceitar uma parte da divisão de Stich com relação a sua visão, pois, em vez de defender só uma das linhas ou negar que essa divisão faça sentido, ele afirma que defende uma *linha flexível*. A linha flexível se comportaria ora como a linha suave e ora como a linha rígida, e que a noção de racionalidade necessitará de um certo ajuste.

Esse ajuste é providencial, pois não é tarefa simples entender a que Dennett exatamente se refere quando fala de racionalidade. O que, então, Dennett fala sobre o conceito de racionalidade que é explorado diretamente pelo estrategista dos sistemas intencionais e explorado como segunda natureza pelo resto da população? A TSI propõe que utilizemos uma pressuposição de racionalidade para a aplicação da postura intencional, mas Dennett constantemente propõe que sistemas reais não serão inteiramente racionais (especialmente se forem resultado da seleção natural) e que devemos levar isso em consideração ao interpretá-los (Dennett, 1978, pp. 11, 21; Dennett, 1987, p. 53). Por isso Dennett avança em suas exigências ao trabalhar teoricamente e recua ao aplicar a postura intencional para casos reais.

Stich is right; for ten years I have hedged and hinted and entertained claims that I have later qualified or retracted. I didn't know what to say, and could see problems everywhere I turned. With that *mea culpa* behind me, I will now take the offensive, however, and give what I think are good reasons for cautiously resisting the demand for a declaration on the nature of rationality while still insisting that an assumption of rationality plays the crucial role I have seen for it. (Dennett, 1987, p. 94)

Como Dennett propõe a sua ofensiva e como ele propõe a definição e explicação do que é racionalidade? Primeiramente, Dennett afirma que não se pode utilizar o fechamento dedutivo e consistência total teórica como critérios necessários para a racionalidade. Ele afirma que quer

usar “racional” como um termo de propósito geral de aprovação cognitiva e os métodos propostos (ou até universalmente aclamados) de se avançar, cognitivamente, no mundo.

Dennett defende a sua definição vaga de racionalidade dizendo que qualquer pessoa poderia se recusar a identificar racionalidade com as características de qualquer sistema formal, ou resultado de qualquer processo, particular e ainda assim realizar apelos ao conceito, e asserções sobre apelos para tal, sem por isso se esquivar de um dever de clareza. Isso acontece porque, como argumenta Dennett, ao investigar-se o conceito de racionalidade, descobre-se que ele é sistematicamente pré-teórico. E quando nos apoiamos no nosso conceito pré-teórico de racionalidade, contamos com as nossas intuições compartilhadas sobre o que faz sentido, pois não há mais nada em que alguém poderia se apoiar.

O argumento principal de Dennett nesse ponto é que a racionalidade é o critério a ser analisado. Quando se perde em racionalidade, inevitavelmente se perde também na capacidade interpretativa ou de atribuição de intencionalidade. Ao mesmo tempo, sempre que houver a possibilidade de uma interpretação intencional, a racionalidade do indivíduo estará lá (o que preencheria uma exigência de necessidade para o critério de racionalidade). Apesar disso, Dennett tem dificuldade em explicitar exatamente com isso funcionaria. Ao argumentar a partir do exemplo citado do menino na barraca de limonada, Dennett mostra a sua dificuldade em provar o seu ponto ao colocar o ônus da prova para seus críticos:

My perhaps tendentious examination of a single example hardly constitutes an argument for my general claim that this will always be the outcome. It is presented as a challenge: try for yourself to tell the total belief story that surrounds such a simple error and see if you do not discover just the quandary I have illustrated. (Dennett, 1987, p. 87)

No final das contas, o que Dennett elabora sobre a natureza da racionalidade utilizada pelo critério na TSI infelizmente não é o suficiente para podermos precisá-la e entendê-la de maneira a ter maior clareza com relação a como abordar os casos de irracionalidade que estão na periferia da nossa capacidade de interpretação intencional. Assim, a dificuldade e inconsistências levantadas pelas formas diferentes de Dennett de encarar a racionalidade em sua própria teoria se mantém firme. Como Dennett mesmo admite, ele não consegue detalhar rigorosamente qual é a natureza da racionalidade e que função ela teria dentro do processo de aplicação da postura intencional. Portanto, mesmo que a distinção levantada por Stich entre a linha suave e a linha rígida não se sustente, ela aponta para uma dificuldade mais profunda da teoria de Dennett em conseguir precisar qual é a função do critério de racionalidade dentro da TSI. Enquanto Dennett não definir com clareza o que ele quer dizer com racionalidade, não só

seus leitores continuarão a ter dificuldade para entender precisamente o que ele quer dizer, como a própria TSI ficará carente de uma explicitação de um de seus principais conceitos, o que dificulta em grande medida a sua exploração e ampliação.

Bortolotti também desafia e critica o critério de racionalidade proposto por Dennett na TSI. Em sua argumentação, ela propõe que há dois conceitos marcadamente distintos de racionalidade dentro da teoria de Dennett e de outros autores que defendem a racionalidade como critério para atribuição de crenças. Assim como a distinção entre a linha suave e a linha rígida de Stich nos auxilia na exploração do conceito de racionalidade como utilizado por Dennett, a distinção das noções de racionalidade e a sua crítica ao critério proposto por Dennett também ajudará nessa investigação.

3.3.2 As duas noções de racionalidade em Bortolotti

Bortolotti identifica essa tensão na teoria de Dennett e propõe que o sucesso da interpretação em alguns casos de irracionalidade não aconteceria, como argumenta Dennett, pelo fato de que a racionalidade de fundo do agente refletiria no seu comportamento possibilitando uma interpretação intencional. O sucesso aconteceria pelo fato de teóricos como Dennett acabarem utilizando dois tipos diferentes de conceitos ambigualmente são chamados pelo mesmo nome. Seriam, todavia, dois sentidos distintos de racionalidade: (1) um critério geral do indivíduo como sistema, perante os diversos critérios normativos de otimização, i.e., adequação às normas da lógica, probabilidade e outras teorias normativas focadas na otimização do comportamento; e (2) como um tipo de racionalidade instrumental que ela chama de *esquema para a razão prática*, voltada para a racionalização dos comportamentos.

The apparent plausibility of the view that explanation and prediction of behaviour in intentional terms rests on the assumption that the subject behaves largely rationally is due to the equivocation between two notions of rationality – rationality as conformity or subscription to epistemic norms, and intelligibility of observed behaviour. The claim that rationality is a precondition for the ascription of beliefs is much more plausible if we have in mind the weaker notion of rationality that is met by any form of intentional behaviour; but the claim that rationality is precondition on the ascription of beliefs is interesting and non-trivial only if the notion of rationality is intended as conformity or subscription to epistemic norms. (Bortolotti, 2010, p. 99)

Bortolotti conclui, assim, que na tentativa de identificar a marca da ação intencional, alguns filósofos argumentam por uma visão de que não pode existir intencionalidade sem racionalidade, mas a evidência deles só apoiaria a afirmação de que o comportamento intencional pode ser racionalizado (Bortolotti, 2004, p. 364), no sentido de buscar-se uma razão,

que seria uma combinação de estados intencionais (crenças, desejos, emoções, etc.) que motivam um sistema a se comportar como o faz (2004, p. 360).

Bortolotti propõe como uma hipótese muito mais plausível a ideia de que quando a explicação e previsão de um comportamento é bem-sucedida é porque o intérprete supõe que o comportamento do indivíduo é inteligível. Não haveria a necessidade de um pressuposto mais forte de racionalidade no sentido de adequação às normas da lógica e probabilidade. Bortolotti propõe que a possibilidade de entender que um garçom ficará desapontado ao receber uma gorjeta baixa e que um garoto estará ansioso por uma visita à loja de brinquedos seja entendido por um apelo a apenas uma noção fraca de racionalidade, uma inteligibilidade do comportamento. Nesta noção existiria a racionalização do comportamento. Bortolotti apresenta um bom exemplo para ilustrar esse tipo de racionalidade:

The schema applies to his action:

1. *Belief*: Stewart believes that it is windy.
2. *Desire*: He wants to be protected from the wind.
- So, 3. *Action*: He puts on his hat before going.

Stewart's action is rational because it makes sense given Stewart's set of beliefs and desires. In other words, (3) is a reasonable conclusion given (1) and (2). Stewart's intentional states (1) and (2) rationalize action (3). (Bortolotti, 2004, p. 363)

Bortolotti também aponta onde a linha de demarcação que separa as duas noções de racionalidade se encontra, ou seja, onde a inteligibilidade do sistema ainda poderia ser encontrada, porém o critério de racionalidade como adesão à otimização do comportamento não estaria mais presente.

Ela comenta que colocar um chapéu, seguindo o exemplo dado, não seria necessariamente a melhor opção disponível para Stewart nas circunstâncias pois poderia falhar em satisfazer outros critérios. Por exemplo, afirma Bortolotti, o chapéu poderia voar se o vento fosse muito forte ou poderia não o aquecer o suficiente para protegê-lo do frio intenso. A questão, portanto, não seria se (3) se seguiria de (2) e (1), mas se (3) é a coisa racional a ser feita naquelas circunstâncias. Essa nova questão apelaria para a racionalidade como otimização, se assemelhando às teorias econômicas e às teorias psicológicas do raciocínio humano (Bortolotti, 2004, p. 363).

Inteligibilidade como elucidação para o critério mínimo de racionalidade

Se for o caso de que a crítica de Bortolotti à utilização indiscriminada por Dennett de duas

noções diferentes de racionalidade se sustente, então a TSI não mostraria exatamente o que Dennett gostaria, ou seja, a pressuposição de racionalidade não seria um critério essencial para a utilização da postura intencional, mas somente um critério mais brando de inteligibilidade, o que deixaria indeterminada a função precisa da racionalidade dentro da interpretação intencional.

Dennett não escreve uma resposta direta às críticas de Bortolotti, mas uma argumentação contrária a essa questão pode ser encontrada em seus textos. A sua defesa consistiria em mostrar que a TSI trata de maneira diferente, sim, os dois tipos de noções de racionalidade—ela apenas não explora objetiva e conscientemente essa distinção. Demonstrando isso, o próximo passo para a sua defesa seria mostrar que os dois modos não seriam contraditórios, mas sim duas formas de uma mesma utilização do critério de racionalidade.

Para começar a elaboração da defesa de Dennett, analiso algumas convergências entre os autores. Apesar dos embates de ideias apontados entre os autores, a posição de Bortolotti quanto à acessibilidade e possibilidade de atribuição de crenças ainda se assemelha muito à posição de Dennett quando são levados em consideração os seus critérios finais para a atribuição de crenças. Para eles, o critério final mais relevante para a possibilidade de atribuição de crenças é a capacidade da atribuição das crenças em prever e explicar o comportamento do sistema.

Then I will argue that any object — or as I shall say, any system — whose behavior is well predicted by this strategy is in the fullest sense of the word a believer. What it is to be a true believer is to be an intentional system, a system whose behavior is reliably and voluminously predictable via the intentional strategy. (Dennett, 1987, p. 5)

I have argued that we are better off if we ascribe beliefs to at least some of the people who suffer from delusions, because their behaviour can often be successfully explained and predicted within the intentional stance. They behave as believers when they act on their delusion, defend the content of their delusion with simple arguments and relate the content of their delusion to other beliefs they hold. (Bortolotti, 2005, p. 206)

Portanto, se eles discordam com relação ao caminho que é necessário para chegar em tal previsão ou explicação, no mínimo eles concordam quanto ao destino final.

Mesmo levando em consideração a semelhança mencionada, ao se investigar mais atentamente os seus escritos, percebe-se que pode haver ainda mais em comum. Para tanto, analiso um exemplo prático de como Bortolotti interpretaria, segundo a sua teoria, uma situação limítrofe de racionalidade. Bortolotti mostra que, mesmo em um caso em que uma pessoa,

Mark, no exemplo, viole claramente um padrão de consistência em suas crenças, acreditando que p e não- p , q e não- q e assim por diante, pode-se, ainda, entender o seu comportamento intencionalmente caso seu comportamento seja guiado por uma crença de cada vez dentre seu par de crenças contraditórias.

I don't think that the project of interpreting people like Mark is doomed. At any one time Mark will endorse either p or not- p and the interpreter will make an attribution on the basis of Mark's behavior at that stage. Why did Mark cut down the tree? Because he believed it was dead. Why was he so upset afterwards? Because at that time he also believed it was alive. Did he change his mind between starting to cut the tree and becoming upset? No, but one belief was more operative at the beginning and the other became operative later. Only if Mark is explicitly inconsistent and endorses the conjunction of two contradictory statements will the interpreter have to wait for other clues in his behavior before attempting an ascription or be resigned to ascribing a pair of inconsistent beliefs. If Mark's adherence to a schema for practical reasoning is totally compromised, then Mark will not count as an intentional system, but the reason for that will be his failure to satisfy the conditions for intentionality and not his failure to satisfy the standard of consistency. (Bortolotti, 2004, p. 365)

Bortolotti afirma que é a falha de Mark em satisfazer as condições para a *intencionalidade* que é responsável pela impossibilidade de compreensão de seus comportamentos e não a falha em satisfazer um critério de *racionalidade* do sistema, a saber, a consistência.

Percebe-se, novamente, a afirmação de Bortolotti sobre as duas noções de racionalidade de otimização e inteligibilidade, cada qual com a sua função, de facilitar a interpretação ou de ser imprescindível para ela aconteça, respectivamente. No entanto, em seu texto, logo em seguida, Bortolotti vai um pouco além da sua tese e argumenta em seguida mesmo uma noção mais trivial de racionalidade, como a inteligibilidade, seria desnecessária para a intencionalidade e somente faria parte de uma heurística de interpretação, algo que facilita o nosso processo interpretativo, mas que não é necessário para que ele aconteça.

Para entender esse passo de seu argumento é necessário compreender o que está envolvido na atribuição de intencionalidade. Para Bortolotti, a atribuição de intencionalidade é uma generalização possibilitada pela PSC que está baseada no conhecimento de fundo (*background knowledge*) do intérprete, aprendida com rapidez, revisável e largamente bem-sucedida (Bortolotti, 2014).

Assim, para a autora, mesmo a sua noção de inteligibilidade é rigorosa demais para restringir a interpretação e atribuição de intencionalidade, pois, servindo apenas como um parâmetro para ajudar a guiar o processo, apenas mostraria uma maneira de alcançar o objetivo final, a interpretação. Resumindo, para Bortolotti, o critério de racionalidade é apenas um processo heurístico que não tem a função de restringir a interpretação, mas apenas de guiá-la.

Para sustentar a sua argumentação, Bortolotti (2004) ressalta que, em alguns contextos,

também adotamos a heurística da similaridade e a heurística da plausibilidade para darmos conta da interpretação. Nenhuma das heurísticas se aplica, afirma ela, para todos os contextos e para todas as instâncias de comportamento que possamos estar interessados em interpretar. Dependendo do contexto, as heurísticas serão adaptadas para o evento, assim como uma pessoa embriagada ou traumatizada pode não agir de maneira racional e também deixar de agir de maneira similar à como nós agiríamos.

Apesar de Bortolotti introduzir a noção de racionalidade por inteligibilidade (ou racionalização) para logo em seguida descartá-la como critério relevante para algum tipo de restrição da interpretação por um critério de racionalidade, as discriminações e os conceitos criados podem ser de grande relevância como ferramenta para auxiliar na análise do uso que Dennett faz do seu critério de racionalidade.

Observa-se que nos exemplos de Bortolotti em que há *inteligibilidade* há sempre um mínimo de racionalidade que é encontrado, algo como um critério mínimo necessário: a necessidade de não-contradição para crenças operantes em um determinado momento. Isso ocorre mesmo quando Bortolotti afirma que só seria problemático que alguém possuísse duas crenças contraditórias se as expressasse no mesmo momento (pois mesmo que as crenças estejam mais ou menos ativas a cada momento, ainda assim uma consistência mínima estaria sempre presente). Não somente essa condição seria necessária para a inteligibilidade, como também o princípio de que crenças, desejos e comportamentos devem cumprir o seu papel funcional para que a noção de intencionalidade e racionalização como proposta por Bortolotti seja satisfeita.

A minha hipótese aqui é de que Dennett vai um pouco mais adiante e afirmaria, em uma ideia da qual Bortolotti discorda, que esses critérios não somente são necessários para satisfazer um tipo de noção de intencionalidade, mas são o *mínimo* necessário para que uma interpretação intencional seja adequadamente realizada. Dennett ressalta a questão do critério mínimo de aplicação para a postura intencional em uma de suas explicações a respeito da aplicação da postura intencional para casos limítrofes. Um exemplo de Dennett tornará isso mais claro. Ele explora o fenômeno da distração ou desatenção (*absent-mindedness*), em que ele poderia acreditar que no seu horário de almoço jogaria tênis e que também almoçaria, dois comportamentos que não poderiam ser conciliados, mas que poderiam estar potencialmente em sua mente caso fossem lembrados. Quando questionado pelo seu parceiro de partida a respeito do motivo de ter ido almoçar, Dennett diz que responderia algo como “Me esqueci completamente disso!” (Dennett, 1987, p. 88).

A única maneira de fazer sentido do seu comportamento é atribuir a Dennett, no

exemplo, dois conjuntos de crenças diferentes em cada momento. Um para quando combinou com o seu parceiro sobre a partida que jogariam mais tarde e outro para quando foi almoçar naquele mesmo horário combinado. Tomados em conjunto temos, portanto, um par de crenças contraditórias que precisariam ser consideradas como não operantes no momento, apesar de ainda presentes, por causa do fenômeno da distração, mas que, para todos os efeitos, funcionam como se a pessoa possuísse crenças diferentes em momentos diferentes.

Da mesma maneira que no exemplo citado por Bortolotti, não houve uma mudança de opinião da pessoa em que algum raciocínio estivesse presente, que pudesse fazer essa mudança de padrão de comportamento parecer válida, compreensível e racional. Como entender uma pessoa que apresenta comportamentos indicando crenças claramente contraditórias sem que houvesse um caminho racional que a tivesse sustentado, ou seja, sem que haja uma mudança de crenças baseada em evidências e inferências lógicas? Por esse motivo apelamos para conceitos que não fazem parte de uma explicação puramente intencional, mas de uma explicação que se vale de conceitos no âmbito da postura de projeto e física: seres humanos tendem a esquecer, e posteriormente lembrar, de certas informações dentro de um modelo de memória, i.e., crenças são ativadas ou desativadas de acordo com o funcionamento particular de nossa memória.

Dennett responde que não há como entender tal lapso na crença do indivíduo sem um mínimo de racionalidade. O proposto por Dennett é, primeiramente, um critério de consistência temporal.

I think a primary motive for my assertion is just to banish the possibility that otherwise would arise: I believe that I am playing tennis at lunch and that I am free to go to lunch as usual. I cannot act on both beliefs at once; whichever I act on, I declare the other to have slipped my mind. (...) It is even possible to have something slip one's mind while one is thinking of it consciously! "Be careful of this pan," I say, "it is very hot" — reaching out and burning myself on the very pan I am warning about. The height of absentmindedness, no doubt, but possible. We would no doubt say something like "You didn't think what you were saying!" — which doesn't mean that the words issued from my mouth as from a zombie, but that if I had believed — really believed — what I was saying, I couldn't have done what I did. (Dennett, 1987, p. 89)

Portanto, algo de racionalidade deve ser preservado, mesmo que seja o que Bortolotti denomina como 'inteligibilidade', para que a postura intencional possa ser minimamente aplicada. O que ditará esse critério? Como foi dito, neste ponto Bortolotti e Dennett se assemelham, pois buscam da mesma fonte a motivação para caracterizar uma interpretação intencional: o sucesso na previsão de comportamentos na utilização de termos intencionais na interpretação de um sistema.

Em suma, Bortolotti sustenta uma visão de que os processos heurísticos de interpretação realizados pelas pessoas são *somente guias* para a interpretação, e não *critérios de restrição*. Assim, considerações de racionalidade ou mesmo plausibilidade e similaridade não precisam determinar se um comportamento tem a possibilidade de ser descrito em termos intencionais, ou seja, se o sistema pode ser descrito como possuidor de crenças ou não (2004, p. 370). Aliás, para Bortolotti, essas considerações nem mesmo *podem* ser critérios que restringem a atribuição de crenças, pois, como ela entende, as práticas da PSC dão conta de explicar e prever comportamentos que não se utilizam nem de critérios de similaridade (como afirma Stich), nem de critérios de racionalidade (como afirma Dennett), mesmo que seja um critério fraco de racionalidade como o designado critério de inteligibilidade.

Vimos que a crítica de que Dennett estaria se utilizando de um critério de racionalidade injustificadamente pelos argumentos apresentados não se sustenta. Vimos também que a dificuldade de Dennett para delimitar exatamente o que ele quer dizer com a utilização do seu conceito de racionalidade gera obstáculos consideráveis para o entendimento da sua teoria e a sua aplicação adequada para a aplicação em casos limítrofes. Dessa maneira, a grande contribuição da noção de inteligibilidade como argumento é que ela nos ajuda a delimitar a fronteira de racionalidade mínima para que um sistema possa ser entendido através da postura intencional, o que já seria um passo considerável nos interesses do presente projeto.

3.4 A contribuição de Dub

A tensão exercida por dois tipos diferentes de critérios de racionalidade dentro da TSI não é apenas percebida e criticada por Stich e Bortolotti. Em uma crítica mais recente, Dub (2015) argumenta contra o critério de racionalidade a partir de duas frentes. Ele explicita duas origens diferentes para as ideias e para os elementos centrais na teoria de Dennett, inferindo que Dennett possui dois princípios distintos que dão suporte para o seu uso do critério de racionalidade. A partir dessa análise, ele faz uma tentativa de refutá-los separadamente e conclui, em concordância com Bortolotti, que a racionalidade do indivíduo poderá, sim, ser algo presente em sua capacidade intelectual, algo que, no entanto, deve ser descoberto, não pressuposto.

Dub identifica a origem das ideias de Dennett em Willard van Orman Quine e Gilbert Ryle, ambos grandes vultos da filosofia do século XX que estiveram nas instituições de ensino nas quais Dennett estudou e serviram como fortes influências e inspirações para este. Para Dub, contudo, nenhuma das influências justifica a pressuposição de uma racionalidade não só no ato

de interpretação, mas como própria possibilidade do indivíduo ter uma mente, assim como faria a TSI. Ele analisa separadamente a maneira como a herança quineana e a herança ryleana influenciaram na teoria de Dennett e conclui que nenhuma delas justifica o seu polêmico pressuposto de racionalidade. Dub propõe, como modo de corrigir as falhas encontradas na teoria de Dennett, uma mudança na TSI de modo a abandonar o pressuposto de racionalidade. Ele sugere transformar a pressuposição *a priori* de racionalidade em uma possível descoberta empírica, salvando o que há de melhor na postura intencional e deixando de lado esse seu ponto de fraqueza teórica.

A argumentação de Dub começa com a crítica padrão direcionada a Dennett, muito similar às encontradas nas discussões anteriores do presente capítulo: não somente há vários casos de óbvia irracionalidade que nós testemunhamos na vida ao lidar uns com os outros, como também a psicologia e a economia demonstram várias maneiras pelas quais somos demasiadamente irracionais.

3.4.1 A linhagem quineana

Dub (2015) identifica o critério de racionalidade no *princípio de caridade* em outros dois discípulos de Quine além de Dennett: Donald Davidson e David Lewis. Segundo Dub, estes autores, ao seguirem os passos de Quine, aplicaram o seu princípio de caridade para outros projetos além da tradução linguística. Para Davidson e Lewis, o princípio de caridade seria um fator limitante que preservaria a racionalidade durante a interpretação radical. Apesar de a interpretação radical ser diferente da tradução radical por atribuir estados mentais, o princípio de caridade ainda operaria de maneira muito similar.

De acordo com Dennett (1987), é necessário reduzir o número de predições possíveis para que a interpretação possa funcionar e para que se possa adquirir algum poder preditivo na interpretação e, segundo Dub, é precisamente por isso que Quine, Davidson e Lewis se apegam tanto ao princípio de caridade. Todos eles concordam que é necessário limitar a interpretação através de um requerimento de racionalidade para que se consiga qualquer poder interpretativo ou preditivo.

Dub distingue, ainda, dois projetos: a busca de atribuição de crenças realizada pela PSC que ele chama de *atribuição individual*; e a busca pelo que é desenvolvido por cientistas e filósofos no desenvolvimento de suas teorias, que ele chama de *atribuição científica*. Dub afirma que os escritos de Dennett buscam o segundo tipo de atribuição, almejando uma teoria

madura da intencionalidade e dos estados mentais, que possibilitaria algum desvio dos termos ordinariamente utilizados pela PSC.

3.4.2 A linhagem ryleana

A segunda forma de Dennett entender o critério de racionalidade, segundo Dub, é a partir da linhagem ryleana. Dub nota que é uma análise conceitual do vocabulário mentalista que leva Dennett à sua segunda versão do requerimento de racionalidade. Ela seria a suposição de que significados de termos mentais são determinados por sua conexão e relação com outros termos mentais. Dessa maneira, explicações intencionais assumem a correta e apropriada conexão entre antecedente e consequente.

Seguindo a mesma forma de análise, para que uma partícula subatômica seja considerada um elétron ela deve ter determinadas características. Deve ter carga negativa; deve ter um momento angular intrínseco de $\frac{1}{2}$; e assim por diante. Se alguma partícula sob observação não demonstrar essas propriedades, simplesmente não é um elétron. De maneira similar, para que um estado mental tenha o papel de uma crença, deve estar em relações causais ou funcionais com outros estados mentais (Dub, 2015, p. 106).

Dub conclui sua distinção dizendo que a afirmação de que crenças são constitutivamente racionais pode ser lida de duas maneiras, sendo que elas não são sempre diferenciadas. Primeiramente, pode-se pensar que o processo de interpretação envolvido na atribuição de estados mentais é limitado por um princípio que garante a racionalidade do agente interpretado. Alternativamente, pode-se entender que é uma característica do papel funcional da crença que seja racional: se um estado mental não servir um papel de uma crença que é formada e mantida racionalmente e que motiva o comportamento de uma maneira racional, então não tem o papel de uma crença. Na primeira tese, quineana, a racionalidade é uma condição para o processo de interpretação, enquanto que na segunda, ryleana, a racionalidade é uma característica necessária das consequências do processo de interpretação.

3.4.3 A argumentação de Dub

Contra a linhagem quineana

Dub argumenta que seria necessário realizar essa distinção para uma discussão adequada do

tema, mas também argumenta que ambos os requerimentos estão equivocados e que uma alteração da TSI transformando a racionalidade de um requerimento em algo a ser potencialmente descoberto nos agentes a serem interpretados é o caminho correto a ser seguido.

Dub chega a essa conclusão questionando o critério de racionalidade com duas perguntas: (1) Seria toda restrição na interpretação uma restrição de racionalidade? (2) Uma restrição de racionalidade é, de todo, necessária? Dub conclui que já existem princípios que limitam a indeterminação e uma restrição de racionalidade adicional não é desejável, nem interessante. O autor afirma que se o único objetivo é reduzir a indeterminação dos estados mentais, não é nada óbvio que a restrição de racionalidade é a única restrição ou pressuposto que deveria realizar essa tarefa. Essa seria uma das opções viáveis, mas além de existirem outros modelos de representação que atingiriam o mesmo objetivo, em determinados momentos, essas outras opções seriam mais viáveis do que um modelo de redução da indeterminação que se apoie somente na racionalidade, pois muitas vezes presumir a irracionalidade é mais preditivo do que o contrário.

Segundo Dub, Dennett afirma que quando um indivíduo não pode ser coerentemente interpretado a alternativa correta seria mudarmos para a postura de projeto ou para a postura física. Mas não haveria motivo para isso enquanto ainda existirem outros modelos da psicologia humana que são similares àqueles aplicados pela postura intencional, pois não é somente devido à nossa capacidade de interpretar seres humanos através da postura intencional que deveríamos parar nela em nossa tentativa de explicá-los. Outros modelos mais preditivos poderiam existir, os quais permitiriam uma atualização da postura intencional. Isso seria justamente o que a psicologia cognitiva contemporânea faz, afirma Dub, quando descobre outros estados mentais.

Segundo Dub, o critério de racionalidade força a interpretação do comportamento humano a uma possível interpretação do comportamento humano que nem sempre será a interpretação mais compreensível e que gera mais previsões. Haveria outros princípios que regulam de maneira adequada a nossa construção teórica e, portanto, não haveria a necessidade de mais um critério para reduzir a indeterminação. O que nos previne de ficarmos presos na indeterminação são os nossos comprometimentos com princípios epistêmicos ou virtudes científicas que já guiam as nossas análises e construções teóricas, como simplicidade, conservadorismo, fecundidade, etc. Nenhuma outra ciência além da psicologia, ressalta Dub, parece precisar de um critério adicional para suas investigações.

Quando são utilizados pressupostos de racionalidade para atribuição de crenças, é verdade que se obtém poder preditivo. Mas o trabalho elucidativo está sendo realizado pela pressuposição de que teorias com um poder preditivo são melhores, não que teorias que

pressupõem racionalidade são melhores. É a virtude epistêmica da previsibilidade e sistematização, continua Dub, que nos impede de atribuir crenças irracionais, não um distinto requerimento de racionalidade. Pode-se notar que a racionalidade não pode ser derivada diretamente da previsibilidade e da sistematização, pois se há uma teoria que prevê quão irracionais serão os seres humanos, a teoria mais preditiva, sistematizada e empiricamente adequada pode ser aquela que interpreta os agentes como sendo irracionais. Mesmo que se presuma que modelos que pressupõem a racionalidade devem ser adotados por serem mais simples do que outros modelos, a força da argumentação está no fato de que teorias mais simples são mais parcimoniosas e não que elas pressupõem mais racionalidade.

Dub conclui que se interpretarmos agentes como racionais porque somos levados a essa interpretação pelas normas científicas de previsibilidade, sistematização e adequação empírica, então a racionalidade não precisa ser um critério restritivo para a interpretação. Poderia ser, em vez de um pressuposto, um resultado da psicologia que agentes seriam amplamente racionais, afirma Dub.

Contra a linhagem ryleana

Apesar de desenvolver bem a sua argumentação contra a linhagem quineana, Dub não se detém muito na elaboração de sua resposta à linhagem ryleana antes de dispensá-la sem grande cerimônia. Para ele a discussão perde o seu sentido quando se modificam os termos da PSC em busca de um modelo da psicologia humana que seja mais interessante. Não haveria motivos para pensar que os termos da PSC são sagrados, e, portanto, se deter aos resultados de uma análise conceitual de termos advindos da própria PSC não levará ninguém a uma teoria madura e mais interessante. Segundo Dub, mesmo que o próprio conceito de intencionalidade da PSC estivesse ligado analiticamente à racionalidade, ainda assim poderíamos e deveríamos modificá-lo em busca de um conceito que nos sirva melhor.

One might think that it is just central to the meaning of ‘intention’ that it implies rational relations to other mental states. If this were true, then amending intention to be intention-like really would be considered a version of eliminativism. To my ears, this sounds like a semantic dispute over what states merit the name ‘intention’. Arguing over whether an irrational intention is an intention does not sound much different to me than arguing whether a wrap is a sandwich (Dub, 2015, pp. 107-8)

Dub conclui a sua argumentação com uma comparação ilustrativa. Ele afirma que poderia acreditar que elétrons têm carga negativa e mesmo assim se diferenciar notavelmente de alguém que acredite em um requerimento de carga negativa para a construção de teorias científicas.

Digamos, continua Dub, que esta pessoa tente me convencer da sua convicção.

He presents me with the following two arguments: firstly, it is an indeterminacy-reducing constraint on theory-building that physics posit a subatomic particle with negative charge; secondly, it is a central part of our current concept that electrons have negative charge, and the concept is too useful and predictive to ever want to give up. (Dub, 2015, p. 108)

Para Dub, o citado “requerimento de carga negativa” é completamente desnecessário. A existência de partícula subatômicas com cargas negativas foi uma descoberta, não uma condição *a priori* da investigação científica.

3.4.4 A resposta de Dennett às críticas levantadas

Com a sua argumentação, Dub avança uma proposta de preservar a maior parte da postura intencional e de dispensar o pressuposto de racionalidade, abrindo, assim, espaço para descobertas empíricas sobre o quão racionais realmente somos. Apesar de ser um ataque a um dos elementos centrais de sua teoria sobre as crenças, a distinção realizada por Dub entre as linhagens quineana e ryleana parece ser apoiada por Dennett. Em sua resposta direta ao texto de Dub, Dennett elogia a maneira pela qual Dub expõe a questão e ressalta que o desacordo entre eles parece surgir principalmente pela diferença de aceitação e apreço pelos insights de Quine e Ryle.

Dub baffled me at first, in a very useful way. How could he know the relevant literature so well (not just my work, but Davidson, Lewis, Quine, Ryle, Stich, ...) and still not “get it”? Since he lays out the issues better than anyone I’ve ever encountered, I conclude that I am probably the one who is missing something. (Dennett, 2015, p. 206)

As Dub shows, both of my mentors, Quine and Ryle, have contributed to my confidence that this is the wise way to proceed, and I speculate that the residue of disagreement and/or misunderstanding between Dub and me is largely due to his not entirely sharing my enthusiasm for their insights. (Dennett, 2015, p. 208)

Dennett aceita a categorização realizada sobre a sua teoria, mas rejeita a conclusão de que o critério de racionalidade seja desnecessário e que abandoná-lo em favor de um critério mais abrangente que permita a descoberta da irracionalidade de organismos, como a redução da indeterminação ou aumento da parcimônia, seja estritamente melhor. Para explicar a sua defesa, Dennett reforça que o poder preditivo adquirido pela PSC advém justamente da sua ousada idealização racional. Idealização racional essa cuja utilidade se estende para interpretações a respeito de computadores, animais ou da própria evolução por seleção natural. Portanto,

diferentemente do que afirma Dub, não seria só na psicologia que o critério de racionalidade seria relevante. Na própria biologia a pressuposição de racionalidade, inerente a perguntas norteadoras como “Para que essa estrutura celular serviria?”, é responsável pelo enorme sucesso na investigação e descobertas científicas sobre as características e funcionalidades dos seres vivos. Da mesma maneira que explicações a partir da postura intencional na biologia funcionam porque a seleção natural é uma excelente projetista de seres vivos para o propósito de sobrevivência e procriação, as explicações a partir da PSC também funcionam porque pessoas (bem como animais, plantas, etc.) foram bem “projetados” para se proteger e promover seus interesses.

Assim como biólogos não precisariam, a princípio, de explicações a partir da postura de projeto para desenvolver as suas pesquisas, a psicologia também não precisaria da postura intencional para explicar o comportamento dos seres humanos, mas Dennett aponta para a necessidade de se reconhecer que a enorme capacidade de predição e explicação advém justamente da racionalidade pressuposta e não de algum outro lugar. Seguindo a linhagem quineana, esse seria o método utilizado pela PSC para reduzir drasticamente a indeterminação no processo de interpretação que funciona porque os seres vivos evoluíram (e o método, ele próprio, também evoluiu e existe hoje porque funciona).

Entretanto, essa argumentação só seria adequada para responder ao que Dub cunha de *atribuição individual*, que busca entender como os seres humanos em suas vidas na interação uns com os outros utilizam o termo crença, sendo a *atribuição científica* que seria o mais importante para a empreitada filosófica em busca de uma teoria madura dos estados intencionais. Nesta última, se buscariam conceitos e teorias que promovessem o que há de mais interessante para melhor desenvolver a empreitada científica.

Dub argumenta que ao separar os projetos de atribuição individual e atribuição científica, deveríamos reconhecer que é perfeitamente aceitável que se modifique a PSC se, com isso, ganharmos poder explicativo e preditivo. Como o próprio Dennett faz isso ao introduzir um novo tipo de atitude proposicional que ele chama de opiniões, Dub considera que o dano já estaria feito e a PSC seria passível de modificação em prol de melhores modelos de explicação. Portanto, não há por que se ater ao critério de racionalidade para uma teoria madura da mente simplesmente pelo fato de que é o encontrado na PSC.

Para minar esse argumento, Dennett se vale de explicações que, segundo Dub, seriam advindas de sua linhagem ryleana. Dennett se mantém firme, reafirmando que para que um estado de uma pessoa possa ser considerado uma crença ele deve ter determinadas características e se relacionar com outros estados mentais de um determinado jeito. Não

chamaríamos de crença, afirma, estados mentais que não tivessem essa propriedade.

Similarly, I am claiming, we discovered that intelligent animals have lots and lots of cohering information that they use to guide (appropriately) their actions. We call this information beliefs. We wouldn't call states of a person beliefs that didn't have this delightful property. Electrons are "by definition" negatively charged and beliefs are "by definition" rationally maintained. (Dennett, 2015, p. 208)

Segundo Dennett (2015), não é o caso de nós já sabermos, em algum nível, o que a crença é, ou de só estarmos procurando uma boa teoria para explicá-la. Para ele, a utilização da postura intencional e de conceitos como crenças e desejos não é mais do que o uso de uma ferramenta para identificarmos padrões existentes em nossos comportamentos. Porém, essa ferramenta, apesar de funcionar para a grande maioria das situações, não dá conta de todos os casos possíveis, como é manifesto em casos limítrofes.

Em algumas de suas passagens, Dennett sugere, no entanto, que a PSC, quando legitimada e purificada pela TSI, é o melhor que podemos fazer em grande parte das facetas de nossas vidas (1987, p. 234). Mas ele também defende que nada impede a busca de novos modos de sistematizar os dados obtidos dos fenômenos relevantes e de derivar previsões confiáveis a partir deles, e entende a própria psicologia cognitiva como engajada nessa empreitada. Dennett afirma se juntar a Dub ao aplaudir esforços recentes de Tamar Gendler (2008), Eric Schwitzgebel (2002, 2012), Andy Egan (2009) e Keith Frankish (2009) por suas propostas inovadoras em buscar categorias que acomodem melhor os dados encontrados do que os estados mentais como crenças, hoje utilizados pela PSC. No entanto, Dennett ressalta que o destino dessas propostas e das novas categorias postuladas—se serão identificadas como nada mais do que as crenças e os desejos da PSC ou se os substituirão por completo—deve ser decidido nos moldes da utilização da TSI, de maneira diplomática ou pedagógica e não mediante uma discussão metafísica.

3.5 Princípios norteadores do critério de racionalidade na TSI

Apesar da resposta de Dennett (2015) mostrar que ele ainda se opõe aos levantamentos de Dub, a sua apresentação das duas linhagens norteadores da TSI ilumina de que modo funciona precisamente o critério de racionalidade e qual é a sua função dentro da teoria intencional de Dennett. Como exposto, Dennett aceita a explanação de Dub a respeito das origens de suas

ideias. Utilizo, portanto, sem maiores ressalvas, a discussão de Dub e a sua distinção entre a maneira quineana e a maneira ryleana de entender o critério de racionalidade para refinar a análise em uma ajuda para formular um modo sistematizado e esclarecido dentro da TSI que possibilite a sua aplicação detalhada a casos controversos de atribuição de crença.

Podemos, agora, responder mais adequadamente ao desafio de Stich proposto a Dennett, a saber, especificar precisamente qual é o critério de racionalidade utilizado pela TSI. Como abordado no presente capítulo, Dennett responde que o critério de racionalidade advém de uma noção pré-teórica de racionalidade, algo que não possui fronteiras fixas determinadas, mas que serviria para o propósito de categorizar quando algo serve para o avanço de determinados interesses. Proponho, a partir da análise realizada, que a racionalidade exerce um papel duplo de (a) preservar a relação entre crenças, desejos e comportamentos de modo a possibilitar uma interpretação intencional; e ao mesmo tempo, de (b) servir como pressuposto que possibilite o máximo de poder preditivo com um mínimo de conhecimento sobre o sistema. Estes papéis seriam dois princípios funcionais para a racionalidade dentro da TSI vindo da herança ryleana e da herança quineana, respectivamente.

3.5.1 O princípio da redução da indeterminação

Dennett mostra a aderência ao princípio de redução da indeterminação para a interpretação em muitos momentos em que foca no poder que a PSC e a TSI têm em conseguir ambiciosas predições de comportamento com poucas informações e baixo entendimento do funcionamento interno do sistema. O foco, em grande medida, é a simples capacidade de gerar uma previsão do seu comportamento a partir de certos dados sobre ele. O princípio da redução da indeterminação surge diretamente da definição e do critério utilizado pela teoria de Dennett para crenças e desejos. Segundo sua teoria, não há nada para além da posse de crenças e desejos do que ser volumosamente previsível pela postura intencional (1987, p. 108).

Juntando-se esse princípio com o princípio da indeterminação da tradução advindo diretamente de Quine (1960), Dennett afirma que é sempre possível, em princípio, que interpretações intencionais rivais empatem em primeiro lugar como a melhor interpretação, deixando em aberto a respeito do que o sistema *realmente* acredita e deseja. Dennett sugere um exemplo sobre a escolha de movimentos dentro de um jogo de xadrez para falar abertamente sobre o poder que a aplicação da postura intencional pode trazer para a redução da indeterminação. Ele afirma que podemos na maioria das jogadas (se não em todas) reduzir

drasticamente o número de lances possíveis ao interpretar o nosso adversário. Podemos até, em alguns momentos, ao utilizarmos a postura intencional, reduzir a indeterminação de interpretação completamente. Essas situações acontecem, afirma Dennett (1987, p. 79), em situações (geralmente mais presentes no final) do jogo de xadrez chamadas de “movimentos forçados”. Nesses movimentos a capacidade de previsão do comportamento do jogador pela postura intencional (e somente pela postura intencional) chega a 100%, pois há somente um movimento que satisfaria o desejo do sistema de não perder, levando em conta o que ele tem de informações (i.e., o que ele acredita) sobre o jogo. Para compreender corretamente o papel da racionalidade nesta previsão, é importante notar que movimentos forçados são forçados somente no sentido de serem “ditados pela razão”, pois um jogador irracional poderia realizar o movimento desnecessário que o faria perder o jogo (ou mesmo um jogador racional poderia realizar esse movimento, desde que tivesse um motivo oculto que vá além de ganhar o jogo).

Porém mesmo em circunstâncias em que não se pode ter certeza do comportamento do sistema sob interpretação, Dennett argumenta, a postura intencional ainda assim consegue diminuir a quantidade de possíveis comportamentos em uma proporção significativa. No caso do xadrez, mesmo no meio do jogo, são raros os momentos em que não se pode cortar as possíveis jogadas do oponente consideravelmente, por exemplo, de quarenta movimentos legais disponíveis para uma lista de meia dúzia de movimentos mais prováveis.

Desse modo, não é necessário que se utilize somente a pressuposição de racionalidade para poder prever o comportamento de um sistema. Qualquer método capaz de gerar previsões acuradas e em número grande o suficiente para a previsão do comportamento adequado do sistema pode ser utilizado. Essa é a defesa de Bortolotti ao encarar o critério de racionalidade como um guia para a interpretação, uma heurística que poderia ser substituída por outras heurísticas que auxiliem o processo.

Dennett ainda defende, no entanto, que não é à toa que a PSC se utiliza do pressuposto de uma idealização racional em suas interpretações. Para Dennett, o truque utilizado intuitivamente pela PSC, e explorado conscientemente pela TSI, de uma pressuposição de racionalidade, funciona porque a teoria da evolução por seleção natural é consistente o suficiente para gerar organismos que possuem processos de formação de crença (e desejos) racionais o suficiente para que a grande maioria de suas crenças sejam crenças verdadeiras. Em outras palavras, a TSI funciona em outras pessoas e em outros seres vivos porque eles evoluíram e o processo evolutivo por seleção natural garante a racionalidade dos organismos.

Dennett não elabora explicitamente o seguinte argumento, mas o deixa implícito em suas asserções sobre o tema. Sem a racionalidade para caracterizar os mecanismos de produção

de crenças, não há como garantir que o processamento de informações ocorra da maneira adequada, produzindo crenças verdadeiras. Se a maior parte, ou mesmo uma parte significativa, das crenças de um sistema não for verdadeira, os comportamentos do sistema não o levarão a satisfazer os seus desejos (ou, pelo menos, não poderão garantir que os satisfarão). Isso acontece, novamente, como uma consequência lógica do modo pelo qual os estados intencionais interagem. Se o sistema acredita que A leva a B e tem o desejo de B, ele realizará A. Porém, se esta crença for falsa, ou seja, não é o caso de que A leva a B, então o comportamento A não satisfará o desejo do sistema. Assim, se os comportamentos do sistema não os levam a realizar os seus desejos, não só esse sistema se torna uma impossibilidade evolutiva, como também não há como descobrir quais seriam esses desejos e ainda muito menos quais seriam as suas crenças. Uma forma de entender isso é comparar com uma equação de três termos. Nela somente é possível achar uma solução determinada quando somente um dos termos é uma variável. Quando há mais do que uma variável, a equação se torna indeterminada ou insolúvel. Se sabemos quais são os comportamentos de um sistema e quais são os seus desejos, podemos deduzir e descobrir¹ quais seriam as suas crenças, ou vice-versa. Mas não há como descobrir quais são os desejos de um sistema se as crenças não são coerentes entre si, mesmo sabendo quais são os seus comportamentos.

A lógica dessa argumentação se faz a partir das afirmações de Dennett de que o sistema necessariamente possui a maioria de suas crenças verdadeiras, mas somente naquilo que mais lhe interessa, ou seja, naquilo que seja pertinente para permitir a realização de seus desejos caso seja correto. Pouco importa para um sistema possuir uma crença incorreta que não é relevante para os seus desejos. Há, nesse sentido, uma lógica estruturada em qualquer utilização da postura intencional que necessita de um pressuposto de racionalidade.

3.5.2 O princípio da conexão holística

O princípio de conexão holística é o termo que escolho para designar o princípio advindo da linhagem ryleana como apontada por Dub. O que leva Dennett a esse princípio para o seu critério de racionalidade é a análise conceitual do vocabulário mentalista pela suposição de que significados de termos mentalistas são determinados pelas suas conexões holísticas com outros termos mentalistas. Por exemplo, Dennett (2015) reforça a relação estreita que a crença tem

¹ Descobrir as crenças e desejos no sentido da postura intencional, i.e., de encontrar atribuições de crenças e desejos que possibilitem uma explicação ou previsão correta de seu comportamento.

com outros estados mentais ao afirmar que não chamaríamos de crença os estados de uma pessoa a não ser que tivessem a distinta propriedade de possuir informações que são usadas para guiar ações e a propriedade de serem racionalmente mantidas.

Isso aconteceria a partir do nível mais básico de análise conceitual que determina a função da crença, desejos, outros estados intencionais e comportamentos.

Take, for example, the belief that it is raining. What behaviour would clinch it that A believes it is raining? No matter what is suggested, it will turn out that this is a clincher demonstrating that A believes it is raining only if we assume that A has some particular purposes or intentions. A's saying 'It is raining' or answering 'Yes' to the question 'Is it raining?' only counts as evidence on the assumption, *inter alia*, that A intends not to deceive us and 'intends' is an Intentional idiom. A's finding a tree or roof to stand under is no more evidence, for it depends on A's intending to stay dry. If ascription of belief always depends on an assumed ascription of intention, the converse holds as well. A's intention to stay dry is not behaviourally demonstrated by his cowering under the tree except on the assumption that he believes it is raining, that he believes that he would get wet if he did not stay under cover, and so forth. (Dennett, 1969, pp. 31-32)

Dennett explica como a própria relação causal e o nosso entendimento dela são implicadas na utilização do vocabulário mental intencional.

In the domain of causal explanation, that a particular event a is followed by another event b is explained by the invocation of some more general causal law, to the effect that all events of type A (which includes a) are followed by events of type B (which includes b), and this law may be explained in turn by being subsumed under or deduced from still wider laws. In Intentional explanation, on the other hand, the sequences of events are so characterized that the occurrence of a particular consequent action is explained by the occurrence of a particular antecedent, say a perception or belief or intention, and there is no room for the question of why this consequent should follow this antecedent, and hence no room for any general law 'explaining' the sequence. (Dennett, 1969, p. 37)

Para Dennett, a sua intenção de sair da sala pela porta garante o comportamento de que ele saia da sala, entre outras coisas, pois caso contrário não seria uma intenção de sair da sala pela porta. A pergunta de por que um deveria se seguir do outro é, como Dennett aponta, supérflua, pois já está indicado na própria definição e correta interpretação dos termos intencionais. Dessa maneira, ressalta-se o fato de que a TSI utiliza os termos 'crença' e 'desejo' não meramente como estados utilizados em uma análise estatística de correlações destes com comportamentos, ou mesmo como causas e efeitos como tradicionalmente entendidos, mas estados intencionais que se relacionam de maneira específica com outras crenças, desejos e comportamentos. Pode-se entender isso mais claramente ao vermos que nem todas as explicações que citam termos intencionais são, por esse motivo, intencionais. Para que uma explicação seja, de fato, intencional, é necessário que se forneça um conjunto de razões para o seu comportamento, um

motivo em termos de suas intenções, desejos e crenças do sistema que justifiquem um determinado comportamento.

Dennett (1978, p. 236) esclarece a diferença com alguns exemplos para diferenciar o sentido de causalidade dos termos intencionais sendo citados. Primeiramente, em um sentido humeano de causa e efeito, exemplos de frases seriam:

- (1) A sua crença de que a arma estava carregada lhe causou um ataque cardíaco.
- (2) O seu desejo compulsivo por vingança lhe causou úlceras.
- (3) O pensamento da sua fuga apertada da cascavel o fez com que ele estremecesse.

Estas frases se contrastam com:

- (4) Ele se jogou no chão por causa da sua crença de que a arma estava carregada.
- (5) Seu desejo obsessivo por vingança o levou a seguir Jonas até Burma.
- (6) Ele se recusou a pegar a cobra porque nesse momento ele pensou em sua fuga apertada da cascavel (Dennett, 1978a, p. 236).

Em (1), (2), e (3) não há nenhuma racionalidade envolvida na ocorrência do ataque cardíaco, da úlcera ou do estremeçamento, i.e., não há uma razão para seus acontecimentos. Pelo mesmo motivo, essas ocorrências também não são comportamentos realizados pelo sistema, no sentido de comportamentos deliberados pelo mesmo. Por outro lado, em (4), (5) e (6) observa-se uma clara ligação intencional baseado em uma racionalidade por trás dos comportamentos.

A distinção fica ainda mais clara quando o mesmo acontecimento é potencialmente explicado pela causalidade humeana e também pela postura intencional, resultando em interpretações com marcantes diferenças, como é o caso de transtornos psicossomáticos. Pode-se entender que o sintoma de alguém que, digamos, tenha ficado com paralisia seja um mero sintoma de determinadas crenças e desejos, como em (1) e (2). Uma outra maneira de entender os seus sintomas seria de que as suas crenças e desejos levaram a pessoa a querer ficar com paralisia, fazendo com que ela ficasse assim deliberadamente. Enquanto a primeira explicação é meramente um sintoma da crença dentro de uma lógica de causalidade fisicalista, a segunda é uma explicação puramente intencional, uma condição que só aconteceu à luz de seus desejos (e.g., que a pessoa passe a ser cuidada pelos seus parentes ou que os leve a se sentirem culpados) e das crenças de qual seria a melhor maneira de realizá-los (e.g., de alguma maneira se paralisar

ou fingir que está com paralisia) (Dennett, 1978, p. 236).

A argumentação levantada pretende mostrar não só como Dennett entende a estreita interrelação dos estados intencionais, mas também como é necessariamente a racionalidade que liga e faz funcionar a interpretação intencional como um todo:

However rational we are, it is the myth of our rational agenthood that structures and organizes our attributions of belief and desire to others and that regulates our own deliberations and investigations. We aspire to rationality, and without the myth of our rationality the concepts of belief and desire would be uprooted. (Dennett, 1987, p. 52)

Uma outra consequência que advém do princípio da conexão holística, estreitamente conectada ao argumento anterior, é a necessidade de um fundo de racionalidade que se traduza em uma rede de crenças coerentes. Para Dennett, nenhuma crença isolada seria suficiente para justificar a sua atribuição, pois mesmo que ela, por si só, justifique um comportamento ao se comparar com um desejo, ainda assim ela estaria em uma situação de total instabilidade por não estar inserida em uma rede de crenças coerentes e consistentes consigo. Essa situação de instabilidade impossibilitaria, portanto, uma atribuição adequada da crença, pois não seria possível prever ou explicar outros comportamentos através dela. Dennett (1975) se utiliza de mais um experimento mental para montar um argumento para provar o seu ponto. Ele sugere que imaginemos um mundo em que a inserção pontual de crenças é possível por meio de um neurocriptógrafo que grava crenças particulares no cérebro das pessoas. Vamos supor, Dennett continua, que inserimos uma crença falsa na mente de Tom: “Eu tenho um irmão mais velho que mora em Cleveland”. Dennett aponta para duas possíveis consequências do experimento em Tom.

Consider the two outcomes. Tom is sitting in a bar and a friend asks, "Do you have any brothers or sisters?" Tom says, "Yes, I have an older brother living in Cleveland." "What's his name?" Now what is going to happen? Tom may say, "Name? Whose name? Oh, my gosh, what was I saying? I don't have an older brother!" Or he may say, "I don't know his name," and when pressed he will deny all knowledge of this brother, and assert things like, "I am an only child and have an older brother living in Cleveland." In neither case has our neurocryptographer succeeded in wiring in a new belief. (Dennett, 1975, pp. 408-9)

A conclusão de Dennett não é que a escrita cerebral é impossível, mas que para a implantação de qualquer crença necessita da conjunta implantação de muitas outras crenças que possam sustentar a coerência lógica e biográfica do sistema. Assim, ao se entender uma crença da maneira de Dennett, atrelando a sua definição ao seu funcionamento dentro de um sistema que possa ter seus comportamentos coerentemente explicados e corretamente preditos, passa a ser uma consequência lógica que crenças só funcionam como crenças caso venham inseridas em

uma rede de crenças coerentes que possam sustentar um fundo de racionalidade.

Um outro argumento relevante para esse princípio é o argumento evolucionista montado por Dennett, abordado na seção anterior, de que a seleção natural garantiria que a maioria das suas crenças de um organismo seria verdadeira e a maioria das suas estratégias amplamente racionais. Porém, diria um crítico, demonstrar a verdade dessa afirmação, mesmo sem nenhuma ressalva, não seria o mesmo que demonstrar que o processo interpretativo *necessita* de uma pressuposição de racionalidade. O que acontece é que a explicação evolutiva justifica mais o surgimento da PSC como abordagem que identifica esse padrão de racionalidade e o explora, do que justifica a necessidade do critério de racionalidade para qualquer situação em que a interpretação se faz pertinente.

Uma possível resposta para mostrar que a racionalidade é necessária seria dizer que esse pressuposto é indispensável, pois *nenhum* outro pressuposto pode nos ajudar a levantar tanto peso, epistemologicamente falando, quanto a pressuposição de racionalidade. Mas, ainda assim, isso seria mais uma descoberta empírica do que uma necessidade analítica que estamos procurando no momento.

Uma outra maneira de responder à questão se utiliza do argumento evolucionista, mas segue um rumo diferente. A argumentação de Dennett com relação ao critério de racionalidade pela teoria evolutiva sustenta a conclusão de que a pressuposição de racionalidade é uma aposta segura para qualquer organismo encontrado hoje, tendo em vista que o processo de evolução garante que os processos que mantêm um organismo vivo se assemelharão com algo que se aproxima de uma racionalidade ideal. Nesta argumentação, a racionalidade não é explorada pelo processo interpretativo intencional, mas pela própria evolução, criando um processo de interpretação que leve em conta a racionalidade. O que Dennett está dizendo com isso não é que é interessante utilizar a racionalidade dentro da TSI, como se fosse uma escolha dentro da teoria, como insinua Dub (2015), mas sim que a TSI funciona para esses organismos porque ela funciona através de um modelo de pressuposição de racionalidade, e de nenhum outro jeito. O critério de racionalidade, portanto, é intrínseco à postura intencional e à PSC, e por isso ela funciona.

3.5.3 União dos princípios para formar o critério de racionalidade

Para responder ao desafio de Stich ainda é preciso entender precisamente como esses dois princípios se unem para exercer a função que a racionalidade desempenha na utilização da

postura intencional. Quais são as fronteiras do exercício da racionalidade dentro da TSI? O que exatamente Dennett quer dizer com racionalidade quando ele afirma que ela é o solo no qual os conceitos de crença e desejo se enraízam? Como já foi explorado, a pressuposição de racionalidade precisa cumprir o papel de um pressuposto que reduza a indeterminação, mas que, ao mesmo tempo, ainda permita a relação entre desejos, crenças e comportamentos e que garanta que as crenças formem uma rede coerente a respeito do que é de maior importância para o sistema.

Utilizo a análise até aqui realizada para formular de que maneira o princípio de indeterminação e da conexão holística se juntam para formar o critério de racionalidade explorado pela teoria de Dennett. Primeiramente, para reduzir a indeterminação do sistema, a racionalidade possibilita, por excelência, a utilização de regras generativas, podendo naturalmente cumprir esse papel. A pressuposição de racionalidade se apresenta como intenso fator redutor da indeterminação, especialmente porque, segundo Dennett, a teoria evolutiva garantiria que os organismos que evoluíram pelo processo de seleção natural teriam a maior parte de suas crenças verdadeiras e maior parte de seus mecanismos de formação de crenças racionais. Outros fatores poderiam ser utilizados para a redução da indeterminação, mas esse é o fator que a TSI explora.

Contudo, outras regras podem ser utilizadas para um maior poder preditivo, particularmente aquelas que são empiricamente aprendidas sobre o comportamento humano ou de outros sistemas intencionais. Assim, é possível admitir regras de produção, manutenção e eliminação de crenças que sejam inferiores à racionalidade ideal por ainda reduzirem consideravelmente a indeterminação, funcionando como incrementos ou exceções às regras mais básicas de racionalidade. Isso é importante, pois, como Dennett (1978) mesmo afirma, uma racionalidade ideal não seria possível porque a evolução, além de garantir um nível de racionalidade para os organismos vivos inteligentes, também garante que a racionalidade apresentada pelo organismo seja abaixo do ideal (Dennett, 1987, p. 51). Isto acontece porque há ganhos reduzidos de adaptação ao ambiente à medida em que o desenvolvimento intelectual de uma espécie chega mais próximo de uma racionalidade ideal devido ao aumento do tempo e do custo energético para o processamento de informações, que representam uma desvantagem adaptativa. Isso significa que a seleção natural sempre favorecerá uma posição com melhor custo-benefício do que o apresentado por uma racionalidade ideal.

Quanto de racionalidade em busca da redução da indeterminação deve ser preservada antes de descartarmos a interpretação? Isso depende da nossa tolerância em busca de previsões acuradas. Quanto maior a nossa tolerância para a incerteza, maior será a nossa tolerância para

o distanciamento da racionalidade ideal. Se ficarmos satisfeitos ao poder prever as jogadas mais prováveis do nosso oponente em um jogo de xadrez, não precisamos da pressuposição de um oponente com uma racionalidade suprema, mas apenas racional o suficiente para impedir jogadas claramente contraproducentes ou prejudiciais.

Porém, o princípio da conexão holística também deve ser respeitado, e, na medida que essas outras regras não se utilizam da racionalidade e da relação entre crenças, desejos e comportamentos, elas restringem a intencionalidade que pode ser corretamente atribuída ao sistema. Há, portanto, uma flexibilidade no quanto de racionalidade atribui-se a um sistema, sendo que as regras adicionais delimitam as fronteiras da racionalidade, mas a intencionalidade (e as crenças, por consequência) só acompanha até onde a racionalidade se estende. Assim como não há um organismo 100% racional, também não há um organismo 100% intencional, ou seja, que possa ter o seu comportamento previsto completamente pela atribuição de crenças, desejos e outros estados mentais intencionais. Não é necessário que todo o sistema seja idealmente racional, pois o critério final de análise é que o sistema seja suficientemente previsível pela postura intencional, o que faz com que a tolerância para a falta de racionalidade do sujeito seja motivada, novamente, pela nossa tolerância em predizer o seu comportamento corretamente em termos intencionais. De maneira similar, Dennett (1978, p. 21) propõe que imaginemos ter o conhecimento que em uma população de pessoas que recém começaram a jogar xadrez, o número médio de movimentos ilegais é três. A ideia seria de que essas pessoas estariam jogando xadrez e que poderíamos prever os seus movimentos segundo as regras do jogo, com exceção de algumas jogadas erráticas, não que elas estariam jogando um jogo muito parecido com xadrez, com a única diferença que há uma tolerância para jogadas ilegais fixada em três. Segundo Dennett, assim como não há como prever as jogadas ilegais segundo as regras do jogo, também não há como prever o comportamento irracional pela postura intencional. O que não significa que essas pessoas não estariam jogando xadrez e que não haveria utilidade em entender as suas jogadas dessa maneira. Porém, essa utilidade diminui à medida que os jogadores realizam mais e mais movimentos ilegais. Quando elas param de jogar xadrez? Isso nos leva à próxima questão.

O mínimo de racionalidade para a aplicação da postura intencional

Quanto de racionalidade deve ser preservado para que o conceito de crença e desejo ainda façam sentido como conceitos operacionais? Primeiramente, há um mínimo de racionalidade que deve ser preservado em qualquer interpretação intencional, o que Bortolotti designa como

inteligibilidade. Essa seria a relação lógica que deve existir entre crenças, desejos e comportamentos, a partir das imposições do meio em que o sistema se encontra. O comportamento racional nesse sentido é aquele que, levando em conta as suas crenças, atinge um estado que avance os seus interesses ou desejos.

Em segundo lugar, há um critério abrangente: o mínimo de racionalidade deve ser o necessário para que o mecanismo de produção de crença possa preservar a verdade na criação de novas crenças para aquilo que é relevante para o sistema, isto é, que garanta que ele satisfaça seus desejos. A medida que crenças falsas aparecem no sistema, aumenta-se a possibilidade de que a rede de crenças fique comprometida. Para que as crenças falsas não corrompam o sistema e impossibilitem uma interpretação intencional é necessário, sim, um background de racionalidade e de crenças verdadeiras, porém somente naquilo que mais importa ao sistema. Não precisamos esperar que ele entenda e produza competentemente crenças verdadeiras sobre coisas ou níveis de resolução do mundo que não são diretamente relevantes aos seus interesses. Assim, não se espera que um rato acredite que exista uma cobra cascavel atrás dele, de três cores diferentes e cinco anéis em sua cauda, mas somente que há um predador perigoso o suficiente para que ele fuja dali.

O modelo flexível de racionalidade em Dennett

Todas essas respostas, em conjunto, permitem flexibilidade suficiente na determinação da função da racionalidade dentro da interpretação intencional para que Dennett esteja mais justificado em se recusar a definir o que é *precisamente* racionalidade (Dennett, 1987, p. 94). Sugiro que a definição de racionalidade seja flexível o suficiente para que que *modelos de racionalidade* possam ser postulados para cada interpretação. São modelos de interpretação que limitam a racionalidade atribuída ao sistema para uma previsão mais correta do seu comportamento, mas que, com isso, diminuem a capacidade em interpretá-lo intencionalmente.

Vemos isso quando Dennett escreve, por exemplo, a respeito da necessidade do critério de pressuposição de racionalidade para a interpretação de macacos do Velho Mundo (*Chlorocebus pygerythrus*) pela postura intencional. Ele foca em sua análise sobre os tipos de previsões corretas que podem ser alcançadas se pressupusermos mais, ou menos, racionalidade aos macacos, como sistemas de primeira, segunda ou terceira ordem.²

² Dennett (1987, p. 243) propõe diferentes gradações de sistemas intencionais caracterizadas pelo número de *ordens* que eles são capazes de apresentar. Um sistema de primeira ordem tem crenças e desejos, mas não possui

The question is empirical. The tactic of adopting the intentional stance is not a matter of replacing empirical investigations with aprioristic ("armchair") investigations, but of using the stance to suggest which brute empirical questions to put to nature. We can test the competing hypotheses by exploiting the rationality assumption of the intentional stance. We can start at either end of the spectrum; either casting about for the depressing sorts of evidence that will demote a creature from a high-order interpretation, or hunting for the delighting sorts of evidence that promote creatures to higher-order interpretations. (Dennett, 1987, p. 247)

Na seguinte passagem, fica evidente uma das maneiras em que a teoria de Dennett possibilita que a racionalidade seja inferior à racionalidade ideal:

The rationality assumption thus provides a way of taking the various hypotheses seriously — seriously enough to test. We expect at the outset that there are bound to be grounds for the verdict that vervet monkeys are believers only in some attenuated way (compared to us human believers). The rationality assumption helps us look for, a measure, the signs of attenuation. (...) Once one gets the knack of using the rationality assumption for leverage, it is easy to generate further telling behaviors to look for in the wild or to provoke in experiments. (Dennett, 1987, p. 249)

Alguns dos exemplos de modelos diferentes de racionalidade podem ser encontradas nas passagens dos escritos de Dennett a seguir. Nesta, presume-se que o modelo de racionalidade varie de acordo com os seus interesses, de que é importante para o sistema não gastar energia demais pensando em todas as possibilidades, pois o desejo final do sistema não é realizar a jogada perfeita agora, mas de realizar uma jogada “boa o suficiente” dentro do custo energético e de tempo disponível:

What is Black not likely to notice? I can tell you without even knowing who Black is. Being approximately rational, Black is not likely to notice threats that would take a great deal of time and effort to discover and is extremely likely to notice obvious threats. If Black is, as Fodor supposes, rather unlikely to notice the threat, it must be because the threat is somewhat distant in the search tree and hence may well fall outside Black's more or less optimal focus of attention. So even when we are planning to exploit another rational agent's foibles, we make use of the rationality assumption to guide our efforts. (Dennett, 1987, p. 80)

Então podemos entender um sistema como intencional se ele apresentar algum tipo de modelo de racionalidade, não necessariamente uma racionalidade otimizada para tudo, mas otimizada somente para o que mais interessa para o sistema.

crenças e desejos *sobre* outras crenças e desejos. Já um sistema de segunda ordem possui crenças e desejos sobre outras crenças e outros desejos, seus e dos outros. Assim, enquanto um sistema de primeira ordem teria uma crença “x acredita que p”, um sistema de segunda ordem teria algo como “x deseja que y acredite que x esteja com fome”. Um sistema de terceira ordem teria estados como “x quer que y acredite que x acredite que ele está sozinho” e assim por diante.

3.6 Conclusão

O objetivo principal do presente capítulo foi a análise das principais críticas e avaliação das possíveis dificuldades da TSI em lidar com casos de irracionalidade, examinando os seus pontos fortes e fracos, de modo a auxiliar e guiar a sua aplicação prática. Para isso, analisei as principais críticas direcionadas à TSI em virtude de sua aparente dificuldade em lidar com casos de falha de racionalidade. Comecei o capítulo analisando a discussão a partir das críticas de Stich, já que Dennett mesmo entende-as como representativas da aparente dificuldade que a sua teoria apresentaria ao lidar com casos de irracionalidade. Mais recentemente, críticas relevantes foram formuladas por Dub e Bortolotti à possibilidade e ao modo da TSI em lidar com casos de irracionalidade.

Ao se analisar as críticas de Stich e Bortolotti ao critério de racionalidade da TSI, que levam em conta a sua dificuldade em lidar com casos de irracionalidade, identifiquei dois grandes grupos de argumentos: aqueles que se baseiam na dissimilaridade entre a PSC e a TSI; e aqueles baseados na inconsistência na aplicação do critério de racionalidade.

A primeira crítica, da dissimilaridade entre a PSC e a TSI, traria consequências severas para a validade da TSI caso se mostrasse correta. Concluí, no entanto, que, apesar de soar à primeira vista como válida, ela não se sustenta perante uma avaliação mais criteriosa e detalhada dos seus passos argumentativos. Para que ela se mostrasse válida, precisaria demonstrar não somente que há uma diferença entre a PSC e a TSI, mas que há fenômenos que são explicados pela PSC por uma interpretação intencional de que a TSI não poderia dar conta. Observa-se, entretanto, que os exemplos oferecidos, que supostamente deveriam demonstrar isso, na verdade expõem que para cada caso de irracionalidade se paga um preço com o aumento na limitação de explicações intencionais. E, caso ainda haja a possibilidade de uma interpretação intencional, ela só é possibilitada por um background de racionalidade possuído pelo sistema, mostrando novamente a necessidade da intencionalidade do sistema ser sustentada por um tipo de racionalidade.

A segunda grande crítica identificada, por outro lado, demonstrou ter certo mérito. Como Dennett mesmo admite, ele não foi capaz de formular precisamente o que seria a racionalidade explorada e pressuposta pela TSI. Ainda assim, Dennett desenvolve bem o assunto em seus textos e traz algumas diretrizes para entender a aplicação deste conceito em sua teoria. Dennett defende a sua posição de maneira interessante, mas a falta de sistematização e especificação do modo de se utilizar a racionalidade como critério ainda é particularmente

prejudicial ao trabalho que me proponho na presente dissertação. Busquei, assim, uma maneira de sistematizar a sua teoria de modo a viabilizar a aplicação direta da postura intencional para casos limítrofes de racionalidade.

Utilizei-me, para esta tarefa, das considerações de Dub sobre as origens das ideias de Dennett na construção da TSI para revelar dois princípios que moldam a TSI e a sua utilização do critério de racionalidade: o princípio da redução da indeterminação e o princípio da conexão holística. Com esses princípios busquei unir várias considerações de Dennett sobre a função do critério de racionalidade, para, desse modo, viabilizar a construção de uma maneira sistematizada de aplicar a postura intencional.

Ao unir os argumentos de Dennett da maneira proposta, nota-se, como visto na defesa de sua teoria, a necessidade da racionalidade para qualquer interpretação intencional e, ainda, uma importante flexibilidade na sua utilização. A necessidade do critério de racionalidade, como proposto por Dennett, se fundamenta em quatro argumentos principais:

(1) É necessária a redução da indeterminação da interpretação e a pressuposição de racionalidade se propõe a essa tarefa, pois a evolução garante a racionalidade de organismos vivos;

(2) Qualquer interpretação deve levar em conta a estreita relação entre crenças, desejos e comportamentos, que é pautada em um mínimo de racionalidade;

(3) A redução da indeterminação não pode ser alcançada sem o pressuposto de racionalidade; e

(4) Para qualquer interpretação e atribuição de crenças e desejos, há a necessidade de uma rede de crenças e desejos sustentada por um background de racionalidade.

Concluí, finalmente, que a flexibilidade da utilização do critério de racionalidade se pauta na possibilidade de interpretar sistemas imperfeitos, não porque há espaço para uma interpretação intencional em casos de irracionalidade, mas porque há espaço para a explicação de um comportamento por meio de posturas diferentes, intencionais, de projeto e físicas. Compreende-se, assim, a recusa de Dennett em oferecer uma definição de racionalidade ao mesmo tempo em que insiste na sua indispensabilidade.

Capítulo 3

4. Aspectos práticos e teóricos da aplicação da TSI

4.1 Introdução

O que pretendo alcançar com o presente capítulo é uma resposta, mesmo que provisória, para a questão da viabilidade da postura intencional na sua aplicação para casos hipotéticos e particulares de pessoas em situação de irracionalidade. Para alcançar este objetivo, buscarei construir um método mais claro, prático e rigoroso para a aplicação da estratégia intencional para, então, testar sua aplicação de modo que o eventual fracasso da aplicação venha a ser evidência das limitações da teoria de Dennett sobre as crenças.

Para avaliar a viabilidade da postura intencional dessa maneira, pretendo me desincumbir de duas tarefas. A primeira trata-se de explorar como se desenvolvem as críticas e as defesas levantadas no capítulo anterior a respeito da TSI quando pensados à luz da aplicação para casos particulares. A segunda trata-se da análise da capacidade da TSI de conseguir alcançar aquilo que se propõe a ser como modelo teórico: uma redução e legitimização de tudo o que é útil na PSC.

Para tanto, lidarei, na primeira parte do capítulo, com os fundamentos do que seria uma maneira sistematizada e estruturada de utilizar a postura intencional para qualquer caso que demande uma interpretação intencional. A partir destes fundamentos, esboço um método estruturado em passos que devem ser seguidos para a devida aplicação da postura. Este não pretende ir além do que já foi proposto dentro da formulação de Dennett para a TSI, mas pretende ser somente uma forma reorganizada e mais estruturada de aplicar o conjunto teórico de suas obras em casos particulares de interpretação. Construirei tal método tendo em vista as críticas levantadas no capítulo anterior e a melhor maneira de entendê-las dentro do próprio processo de aplicação da postura intencional. Terminarei a primeira tarefa de análise da viabilidade da TSI lidando com as questões teóricas enfrentadas no segundo capítulo que agora podem ter alguns de seus pontos mais facilmente discutidos.

Em seguida, aplicarei o método formulado para diversos casos particulares distintos para entender as dimensões das potencialidades e das limitações da TSI. O primeiro é um caso paradigmático de aplicação da postura intencional para uma pessoa que não apresenta prejuízos em suas funções cognitivas e também não está em uma situação de clara irracionalidade. O

segundo caso é uma situação comum de falha de raciocínio em uma pessoa sem prejuízos significativos em suas funções cognitivas, mas que, mesmo assim, apresenta uma instância de irracionalidade. E, finalmente, analiso o caso de uma pessoa em um quadro psicótico diagnosticado como delírio de Capgras.

Com isso, pretendo aferir a viabilidade da aplicação da TSI a partir da análise da aplicação aos referidos casos; avaliar a capacidade que a utilização da postura intencional alcançou para descrever adequadamente os casos analisados observando a precisão das suas previsões comportamentais, seu poder explicativo e a sua utilidade; e, finalmente, determinar se a utilização foi adequada para o objetivo da TSI, qual seja, explicar tudo o que a PSC explica, mas de um modo formal e sistematizado.

4.2 Bases para a formulação de um método de aplicação

Dennett não elabora um método estruturado e refinado para a aplicação da postura intencional, o que não seria necessário se seu propósito consistir somente na ilustração do funcionamento de certos aspectos da sua teoria. Porém, para aplicá-la a casos limítrofes que a sua teoria deveria dar conta de explicar, faz-se necessária a elucidação e estruturação mais completa de um método. Pretendo realizá-la na primeira parte deste capítulo analisando, primeiramente, o que já foi exposto pelo autor. Dennett fala a respeito dos aspectos da aplicação da postura intencional diversas vezes em seus textos, o que resulta em um delineamento rudimentar da aplicação da postura intencional. A forma mais explícita desta formulação em seus textos se encontra em ‘Three kinds of intentional psychology’.

We approach each other as intentional systems (Dennett 1971), that is, as entities whose behavior can be predicted by the method of attributing beliefs, desires, and rational acumen according to the following rough and ready principles:

(1) A system's beliefs are those it ought to have, given its perceptual capacities, its epistemic needs, and its biography. Thus, in general, its beliefs are both true and relevant to its life, and when false beliefs are attributed, special stories must be told to explain how the error resulted from the presence of features in the environment that are deceptive relative to the perceptual capacities of the system.

(2) A system's desires are those it ought to have, given its biological needs and the most practicable means of satisfying them. Thus, intentional systems desire survival and procreation, and hence desire food, security, health, sex, wealth, power, influence, and so forth, and also whatever local arrangements tend (in their eyes — given their beliefs) to further these ends in appropriate measure. Again, "abnormal" desires are attributable if special stories can be told.

(3) A system's behavior will consist of those acts that it would be rational for an agent with those beliefs and desires to perform. (Dennett, 1987, p. 49)

Neste modelo, buscam-se as crenças que um sistema deveria ter; em seguida se buscam os desejos; e, por final, os comportamentos que seria racional que o sistema desempenhasse.

Consideremos um caso em que esses princípios possam ser aplicados. Em um dia chuvoso, ao observar uma pessoa sair para a rua com um guarda-chuva em sua mão, (1) lhe atribuímos a crença de que está chovendo, pois é o que as suas capacidades perceptivas lhe proporcionam; (2) lhe atribuímos o desejo de não se molhar ao andar pela rua, pois é um meio de alcançar a sua necessidade biológica de não sofrer e passar frio; e (3) prevemos que o seu comportamento consistirá em abrir o guarda-chuva o quanto antes, pois esta seria a ação racional levando em consideração suas crenças e desejos. Esta seria uma aplicação básica do que foi brevemente formulado por Dennett.

Há, ainda, a sua afirmação sobre a atribuição de desejos ou crenças “anormais”. Um exemplo disso seria uma situação em que vemos esta mesma pessoa andando pela chuva, mas sem abrir o seu guarda-chuva. Nesta situação, podemos atribuir-lhe a crença falsa de que ela não teria um guarda-chuva em suas mãos. Para que essa atribuição seja justificada, um tipo especial de história deveria ser contado. Por exemplo, a pessoa poderia acreditar que não segura um guarda-chuva porque talvez ela estivesse tão preocupada em meio a uma ligação telefônica a respeito do estado de saúde de seu filho que nem tivesse notado que tem um guarda-chuva em suas mãos. Outras explicações anormais poderiam satisfazer esta condição, como estar sob efeito de uma indução hipnótica para acreditar que se estaria segurando um buquê de flores, ou ainda como desejar se molhar para lembrar dos banhos de chuva de sua infância. A ideia que Dennett quer passar é que se atribuem crenças verdadeiras às pessoas ou sistemas em questão a não ser que uma história particular possa ser contada para justificar a atribuição de crenças falsas.

Vamos tentar entender como a aplicação da postura intencional ocorre com casos mais complicados em que há uma clara atitude irracional do sistema em questão. Em um estudo clássico de Amos Tversky e Daniel Kahneman (1983), participantes de uma pesquisa foram providos com descrições de uma pessoa e, então, requisitados a avaliarem a probabilidade de algumas afirmações a respeito dessa pessoa serem verdadeiras. A descrição da pessoa era “Linda tem 31 anos, é desinibida e bem inteligente. Ela se formou em filosofia. Como estudante ela era profundamente interessada em assuntos de discriminação e justiça social e participava de manifestações contra a energia nuclear”. Pediu-se aos participantes que avaliassem a probabilidade de certas afirmações, dentre elas: (1) Linda trabalha como caixa de banco; e (2) Linda trabalha como caixa de banco e é ativa no movimento feminista. No estudo original, 85% dos participantes julgaram (1) como sendo menos provável do que (2), violando a regra da

conjunção, que afirma que a intersecção de dois conjuntos não pode ser maior que qualquer um dos dois conjuntos.

Um aspecto importante que deve ser ressaltado é que Dennett em seus escritos se concentra no poder preditivo da sua teoria intencional, mas, junto à capacidade preditiva, viria a capacidade *explicativa* da teoria para comportamentos passados (Dennett, 1983, p. 351; 1994, p. 24). Apesar de elas virem juntas, é necessário mudar um pouco a abordagem para a aplicação da postura intencional quando a utilizamos para explicar o comportamento em vez de prevê-lo, como no caso desse estudo. Desse modo, em vez de questionar-se qual comportamento seria racional o sistema desempenhar dadas as suas crenças e desejos, agora pergunta-se quais seriam as crenças e desejos que racionalmente justificariam esse comportamento (Dennett, 1987, p. 100).

A dificuldade da TSI em lidar com essa situação seria explicar o comportamento tido como irracional, em que os participantes violam a regra da conjunção. Examinemos como seria a aplicação do método em passos proposto por Dennett. O comportamento a ser analisado é a escolha de que seria mais provável que Linda fosse uma caixa de banco e feminista do que uma caixa de banco. Aplicando os passos propostos (1), (2) e (3) temos:

1. Atribuir as crenças que o sistema deveria possuir. Os participantes deveriam acreditar na verdade daquilo que as suas capacidades perceptivas possibilitaram que ele entrasse em contato. O participante deveria acreditar que está diante de uma pesquisa que tem esse questionamento, que é um estudante universitário dentro de um campus, que cursa tais matérias, etc.
2. Atribuir os desejos que o sistema deveria ter. Ele deveria buscar segurança, alimentação, etc. E, relevante para a ocasião, ele deveria desejar responder adequadamente e acertar a resposta da questão indagada na pesquisa.
3. O seu comportamento previsto seria aquele que esperaríamos de alguém que fosse racional frente a esses desejos e essas crenças. Nestas crenças pressupostas não há crenças que definam uma escolha particular dentro do questionário, mas exclui-se um comportamento que seja claramente irracional, como escolher que seja mais provável que Linda seja uma caixa e feminista do que se fosse uma caixa de banco.

O comportamento previsto, no entanto, é justamente o que *não* ocorreu para 85% dos participantes. Parece, portanto, que a TSI deixaria de fazer sentido os resultados da pesquisa. Todavia, ainda há um modo que Dennett propõe para lidar com situações de complicada

atribuição como esta: a atribuição de crenças e desejos “anormais” ao sistema. Será que essas crenças e esses desejos “anormais” poderiam salvar a explicação a partir da postura intencional?

Poder-se-ia atribuir um desejo para a maioria dos participantes de mentir na pesquisa ou responder incorretamente. Dennett parece estar correto em dizer que tal atribuição demandaria uma história especial, por exemplo, que os alunos participantes tenham propositalmente respondido incorretamente a esta pesquisa como uma forma de protesto e retaliação por condutas tomadas como discriminatórias dos professores responsáveis pela pesquisa.

Uma outra saída para essa dificuldade seria a aplicação da orientação de Dennett sobre a racionalidade atribuída a um sistema nestas condições: começa-se com a pressuposição de racionalidade perfeita para o sistema, mas ela é revisada à medida que as circunstâncias se impõem, ou seja, à medida que os resultados obtidos são insatisfatórios (Dennett, 1987, p. 21).

A crítica do *critério duplo de racionalidade* se apresenta com força neste momento. Dennett reitera em diversos textos que um dos pilares da TSI é que o sistema deve ser tratado como *idealmente racional* para que possa ser interpretado pela postura intencional, mas, para salvar a sua interpretação do fracasso quando as circunstâncias não são favoráveis, ele nos orienta a atribuir crenças irracionais ao sistema. Os seus críticos afirmariam que não se pode, de modo *ad hoc*, alterar a sua exigência sobre a racionalidade à medida que resultados negativos retornam da aplicação da teoria, especialmente quando essa alteração é a respeito de um dos pressupostos mais básicos da TSI. Essa inconstância e variação do critério geraria uma tensão que revelaria uma inconsistência teórica grave no seio da TSI.

Já analisei as implicações e as defesas de Dennett para a TSI sobre essa crítica do duplo critério de racionalidade, mas as retomo neste momento para mostrar que ainda não se possui total compreensão a respeito da maneira que a TSI pode ser aplicada para casos de evidente irracionalidade sem levantar questionamentos sobre a sua validade. Portanto, continuo a exploração detalhada da sua teoria em busca de clareza sobre sua aplicação.

Como vimos no capítulo anterior, a presente crítica tem certo mérito, especialmente quando se leva em consideração que o próprio Dennett admite não estar satisfeito com qualquer definição de racionalidade que se proponha fazer jus ao seu uso dentro da sua teoria. Com o intuito de construir um método de aplicação da TSI que lide explicitamente com as críticas levantadas contra o seu critério de racionalidade de modo que as aparentes contradições fiquem evidentes, utilizo-me dos dois princípios propostos e elaborados no capítulo anterior que regem a estrutura interna da TSI, o *princípio da conexão holística* e o *princípio da redução da indeterminação*, para a formulação de dois níveis particulares de aplicação

O primeiro nível de aplicação é o nível de atribuição de crenças e desejos ao sistema, aquilo que Dennett aponta como o que o sistema *deveria* acreditar e desejar, nos passos (1) e (2). Já o segundo nível é aquele em que se busca a previsão de comportamento do sistema e a sua explicação, indicados no passo (3), onde Dennett determina que a atribuição de racionalidade instrumental deve operar. A chave para entender como a separação da aplicação da postura intencional em dois níveis ajudaria a salvar a TSI da contradição é atentar-se para a diferente função delegada à racionalidade em cada nível. Em ambas, o critério de racionalidade cumprirá um papel relevante, mas entender as suas particularidades permitirá atravessar as barreiras restantes para a devida compreensão da aplicação da TSI.

Apesar de os elementos para a presente formulação estarem todos delineados nos textos de Dennett, esta elaboração ainda se faz necessária porque, infelizmente, não há uma formulação e sistematização explícita do autor sobre alguns aspectos do funcionamento da sua teoria, o que facilita interpretações equivocadas e o aparecimento de críticas. Esta divisão proposta, portanto, não é de toda criação minha e ela pode ser vista implicitamente por todo o seu texto, e visivelmente em algumas passagens particulares como a supracitada (Dennett, 1987, p. 49).

Nas próximas seções elaborarei detalhadamente os níveis de aplicação propostos com o objetivo de refinar o modelo de aplicação da postura intencional. Para realizar esta tarefa, utilizo-me das formulações desenvolvidas no capítulo anterior para analisar as particularidades de cada um dos níveis e as suas relações com a racionalidade e a interpretação intencional. Chamo o primeiro nível de aplicação de *nível de atribuição* e o segundo nível de *nível de previsão*.

4.2.1 O nível de previsão

Primeiramente, nesta seção, elaborarei o *nível de previsão*, que é uma abreviação para o nível de previsão e explicação do comportamento. Este, na verdade, é o segundo nível a ser utilizado na execução da aplicação da postura intencional, mas o elaboro por primeiro, pois o seu entendimento inicial facilita a exposição da teoria como um todo.

As diretrizes que Dennett dá para a aplicação deste nível, em sua maior parte, não são explícitas, e, quando são, não são detalhadas. Há somente uma orientação de aplicação de um raciocínio prático para as crenças e desejos atribuídos aos sistemas, como na seguinte passagem.

Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. *A little practical reasoning from the chosen set of beliefs and desires will in many — but not all — instances yield a decision about what the agent ought to do; that is what you predict the agent will do.* (Dennett, 1987, p. 17, ênfase minha)

A noção de racionalidade como inteligibilidade

Por isso, retomo do segundo capítulo uma distinção empregada por Bortolotti, a saber, entre diferentes noções de racionalidade. A noção fraca é chamada de *inteligibilidade* e a noção forte é chamada de *otimização*. Bortolotti (2004, 2010) define a noção de racionalidade por inteligibilidade como a adesão à racionalidade como razão prática. Desse modo, o conceito que ela elabora parece ser justamente o que estou buscando para este nível de aplicação.

A fim de auxiliar no melhor entendimento deste nível de aplicação, analiso os escritos de Bortolotti em que ela discorre sobre a noção de inteligibilidade. Apesar de se deter neste conceito em alguns momentos do seu texto, Bortolotti infelizmente não desenvolve tanto esta noção de racionalidade por não considerar de todo importante, afirmando que o reconhecimento de sua existência seria algo trivial.

The claim that rationality is a precondition for the ascription of beliefs is much more plausible if we have in mind the weaker notion of rationality that is met by any form of intentional behaviour; but the claim that rationality is precondition on the ascription of beliefs is interesting and non-trivial only if the notion of rationality is intended as conformity or subscription to epistemic norms. (Bortolotti, 2010, p. 99)

Note-se, contudo, que toda a discussão é mal-entendida em torno desse tópico e o papel fundamental que a racionalidade tem no entendimento da intencionalidade faz com que pouco do que esteja sendo discutido seja verdadeiramente trivial. O que quero demonstrar é que mesmo na noção mais fraca de racionalidade os meios que utilizamos para dar sentido a qualquer explicação intencional não são tão óbvios ou triviais quanto parecem.

Como a noção de inteligibilidade nos ajudaria a entender o nível de previsão do comportamento? Analisemos novamente o exemplo proposto por Bortolotti (2004, p.363) que representa a noção racionalidade por inteligibilidade:

- *Crença*: Stewart acredita que está ventando.

- *Desejo*: Ele quer se proteger do frio.
- *Ação*: Ele coloca o seu chapéu antes de sair.

O comportamento parece advir direta e logicamente da crença e do desejo, um caso paradigmático da aplicação de uma interpretação intencional. Uma análise cuidadosa, todavia, pode revelar maiores complexidades. Em primeiro lugar, não presumiríamos que o comportamento de Stewart fosse esse se também não lhe fossem atribuídas outras crenças como, por exemplo:

- Stewart acredita que quando ele sair de casa, enfrentará o vento e isto fará com que ele sinta frio.
- Stewart acredita que se ele colocar este chapéu, ele se protegerá do frio ao sair de casa

Na verdade, podemos pensar, ainda, em oito crenças diferentes em que todas são necessárias para a correta previsão deste simples comportamento.

- Stewart acredita que está ventando fora de casa.
- Stewart acredita que por onde se venta faz frio.
- Stewart acredita que ao sair de casa enfrentará o vento.
- Stewart acredita que o vento, ao passar por ele, fará com que ele sinta frio.
- Stewart acredita que chapéus podem proteger uma pessoa do frio.
- Stewart acredita que se ele colocar um chapéu pode se proteger do frio que o vento lhe causaria.
- Stewart acredita que há um chapéu na frente dele que ele pode usar.
- Stewart acredita que se ele colocar esse chapéu em sua frente ele se protegerá do frio ao sair de casa.

O importante a ser ressaltado aqui é que se alguma dessas crenças não fosse atribuída a Stewart, não esperaríamos que ele desempenhasse a ação de colocar o chapéu antes de sair. Isso deve ser evidente, dado que se ele não acreditasse, por exemplo, que o vento passaria por ele (caso ele fosse de um estacionamento fechado até outro estacionamento fechado dentro do seu carro, por exemplo), ele não teria motivo para pegar o chapéu. Do mesmo jeito, se ele não acreditasse que o vento fosse lhe causar a sensação de frio, ele não evitaria o vento com um chapéu, pois

isso somente aumentaria o seu calor se, por exemplo, ele estivesse no verão e sentindo muito calor. Ainda, se ele não acreditasse que o chapéu lhe protegeria do frio, então pegar o seu chapéu para se proteger do frio dificilmente seria uma atitude inteligível. A situação se repete para todas as outras crenças. Assim, para que se chegue à conclusão de que Stewart pegará o seu chapéu antes de sair de casa, seria necessário atribuir (pelo menos) mais sete crenças além da proposta por Bortolotti para que se pudesse fazer uma ligação entre o desejo de Stewart de se proteger do frio e a sua ação de pegar o chapéu antes de sair de casa.

As crenças atribuídas, no entanto, não são gratuitas. Elas devem ser atribuídas a partir do conhecimento de Stewart, do seu contexto e da sua história, sem os quais elas não seriam possíveis ou devidas. Qualquer outra pessoa, no lugar de Stewart, poderia não possuir qualquer uma dessas outras crenças, inviabilizando a inferência realizada.

Outro ponto importante é notar que a ação só será efetivamente explicada intencionalmente no momento em que fizermos uma conexão entre o desejo e a possibilidade do comportamento. No exemplo, a ação de pegar o chapéu e vesti-lo só pode ser explicada, de verdade, no momento em que ligamos o desejo de não sentir frio à crença de que vestir o chapéu fará com que ele não sinta frio. Qualquer falha em um elo fará com que a corrente se quebre, isto é, que a conexão não seja feita e que o comportamento perca o seu sentido.

Para entender claramente quando esta conexão é realizada e quando ela não é, proponho formalizar a referida relação intencional necessária entre crença, comportamento e desejo. O que esta formalização tem que contemplar são os elementos na seguinte relação: a crença de que vestir o chapéu lhe protegerá do frio em conjunto com o desejo de se proteger do frio fará com que ele realize a ação de vestir o chapéu. Em termos gerais isso formaria uma equação da seguinte forma:

- Crença ($A \rightarrow B$) + Desejo (B) = Comportamento (A)

Vejam os a diferença adaptando esta equação para o exemplo proposto por Bortolotti:

- Crença (“está ventando”) + Desejo (proteger-se do frio) = Comportamento (colocar o chapéu)

Percebe-se que os termos propostos por Bortolotti não satisfazem a equação apropriadamente. Para satisfazê-la com o seu exemplo, seria necessário algo mais próximo dos seguintes termos:

- Crença (“colocar o chapéu me protegerá do frio”) + Desejo (proteger-se do frio)
= Comportamento (colocar o chapéu)

O valor desta equação é que ela torna explícito quais são as exigências para uma relação intencional satisfatória. Para tornar esta equação mais precisa, no entanto, é necessário colocar também um outro termo que possa explicar a ausência de ação em momentos em que o sistema acredita que A leva a B, deseja B, e, mesmo assim, não realiza A, porque não dispõe desta capacidade. Dessa maneira, é necessário levar em consideração as *limitações* do sistema quando considera-se o seu comportamento intencional (Dennett, 1978a, p.6). Para tanto, suplemento a equação com um termo que afirma explicitamente a capacidade do sistema em realizar A. Com estas considerações, a equação pode ser reformulada da seguinte maneira:

- Crença ($A \rightarrow B$) + Desejo (B) + Capacidade (A) = Comportamento (A)

Esta é a conexão necessária entre crença, desejo e comportamento formalizada, de modo a explicitar a função principal das crenças e dos desejos dentro da aplicação da postura intencional e motivar o comportamento intencionalmente. Chamo esta equação daqui em diante de *equação intencional*.

A despeito das análises realizadas, há ainda uma questão carente de desenvolvimento com respeito à noção de inteligibilidade elaborada por Bortolotti. Considero que a autora não somente negligenciou as atribuições de crenças implícitas necessárias para estabelecer a conexão entre desejo e comportamento, como também deixou de lado a necessidade da aplicação de um modelo específico de racionalidade.

Para que a previsão do comportamento faça sentido, não só é necessário que sejam atribuídas as crenças citadas, como também há a necessidade de aplicação de algo semelhante a um modelo de racionalidade que possa ser aplicado às proposições acreditadas pelo sistema. Consideremos o exemplo novamente. Deve haver algo que justifique a passagem de 1, 1.1 e 1.2 para 2.

- 1) Stewart acredita que está ventando fora de casa.
 - 1.1) Stewart acredita que por onde se venta faz frio.
 - 1.2) Stewart acredita que ao sair de casa estará onde o vento sopra.
- 2) Stewart acredita que o vento, ao passar por ele, fará com que ele sinta frio.

Bortolotti separa a noção de racionalidade por otimização da noção de inteligibilidade, pois a noção de racionalidade por otimização se utiliza das normas da lógica e da probabilidade que são complexas e buscam nada menos do que a melhor maneira de se chegar a conclusões e inferências corretas. Mas, ao analisarmos de perto a aplicação da noção de racionalidade por inteligibilidade, observamos que a adesão a certas regras lógicas é necessária e não é tão óbvio que possamos atribuir a correta utilização de *modus ponens*, por exemplo, mas que a adesão a outras regras de probabilidade, como a regra da conjunção (como no caso exposto de Linda, a caixa de banco feminista), não. Não há, portanto, algo evidente que delimite que tipo de processo inferencial pode ser atribuído ao sistema e que tipo não pode. Analisemos um outro argumento para elucidar esta conclusão.

Fodor e Lepore (1993, p. 77), em uma crítica à teoria de Dennett, afirmam que um critério de racionalidade ideal como o fechamento dedutivo é excessivo, pois há algum valor preditivo em atribuições de crenças mesmo que uma criatura realize *modus ponens* adequadamente (acreditar que ‘ $p \rightarrow q$ ’, acreditar que ‘ p ’ e, logo, acreditar que ‘ q ’) somente, por exemplo, 87% das vezes. Dennett (1993a, p. 217) responde concedendo que esta afirmação é aparentemente plausível, mas só aparentemente, pois problemas surgem tão logo tentemos nos aprofundar na questão. Digamos ser verdade, sugere Dennett, que somente em 87% das vezes em que uma criatura realiza *modus ponens* ela fizesse isso corretamente, e que, desse modo, em 13% das vezes ela acredita que ‘ p ’, acredita que ‘ $p \rightarrow q$ ’, mas não acredita que ‘ q ’. O problema aparece da seguinte forma: em primeiro lugar, como descobrir que ela tem precisamente esta crença ‘ p ’ e ‘ $p \rightarrow q$ ’ e não outras? Não seria a melhor evidência que poderíamos ter de que um sistema não acredita que ‘ $p \rightarrow q$ ’ de que ele ao mesmo tempo acreditasse que ‘ p ’ e ‘ $\neg q$ ’? Dennett afirma que esta seria não só a melhor forma, mas esse tipo de evidência seria a única evidência que se poderia ter a respeito das crenças de um sistema. Segundo Dennett, qualquer intérprete ficaria relutante em atribuir a crença de que ‘ $p \rightarrow q$ ’ e de que ‘ p ’ a um sistema, se ele apresentasse disposições para acreditar que ‘ $\neg q$ ’.

Que outra evidência poderia demonstrar a crença em certa proposição? Dennett responde que não há outra evidência que possa claramente mostrar a crença em algo. Para ele, esta é uma pergunta que só pode ser respondida com alusão à característica holística da atribuição de crenças, algo que precisa ser explicado com o *círculo Quineano*.

But this assumes that determining, say, that a particular person didn't believe q wouldn't *ipso facto* undo or at least put into question one's earlier finding that the person believed p and $p \rightarrow q$. Sellars, Quine, Davidson, and I (among others) have always seen — or so we thought! — that individual belief attributions were not anchorable in this imagined way because one always encounters what we might call

the Quinian circle: the very best evidence there could be that you had made a mistake in attributing to someone the beliefs that p and $p \rightarrow q$ was the evidence that you should also attribute the belief that not- q . (Dennett, 1993a, 217)

Fodor e Lepore (1993, p. 78) avançam a ideia da existência de leis intencionais: leis que afirmassem que se um estado intencional A é nomologicamente suficiente para que um sujeito esteja em um estado intencional (e/ou comportamental) B, então, dado o conhecimento de que a criatura está no estado A, poder-se-ia prever que ela estará no estado B, quer a transição entre A e B seja racional, quer ela não seja. Dennett concede que se algo como uma lei intencional realmente existisse, esta descoberta teria a força suficiente para romper o círculo Quineano (Dennett, 1993a, p. 218). Mas ele argumenta que esse tipo de lei simplesmente não teria espaço dentro do funcionamento da postura intencional, pois seria algo que estaria funcionando no nível de explicação da postura de projeto ou da postura física.³

Dennett esclarece essa conclusão respondendo ao que seria um suposto caso de lei intencional dado por seus críticos: o tamanho aparente da lua muda com relação a sua posição no céu. Quanto mais perto do horizonte, maior parece a lua para quem a vê da Terra, ou seja, haveria uma relação do tipo encontrada em leis que regeria a relação das crenças perceptivas das pessoas que olham para a lua e a sua posição no céu, sem que houvesse qualquer necessidade de apelos à racionalidade.

Segundo Dennett, a explicação dessa relação seria algo que estaria no nível de explicação da postura de projeto, pois não há nenhum mistério que pode-se haver uma implementação em nível de projeto de competências intencionais, e parece claro que esta “lei” tem uma explicação que diz respeito à funcionalidade das nossas capacidades perceptivas. No final, novamente, para explicá-las como competências intencionais, ligando crenças, desejos e comportamentos intencionalmente, é preciso a pressuposição de racionalidade. Qualquer outro tipo de lei para a criação destes conceitos intencionais teria o formato, dentro da teoria de Dennett, de uma explicação da alçada somente dos níveis de projeto ou físico.

Desta forma, para Dennett não há uma saída isolada para o problema do círculo Quineano. A resposta está na relação holística inerente às atribuições de crenças: não se pode atribuir uma única crença sem a pressuposição de um conjunto de outras, racionalmente relacionadas a esta. Esta é a rede de crenças quineana explicitada pelo *princípio da conexão holística* que deve permear qualquer aplicação da postura intencional.

³ A ideia de que estados intencionais podem ser evocados para explicações de comportamentos e até de outros estados intencionais e que, nem por isso, constituem explicações intencionais foi abordada na seção 2.4.2 e será mais desenvolvida na seção 3.2, mas retomo-a para esclarecer o argumento atual.

Retomei esta discussão para mostrar que ignorar a utilização de um modelo de racionalidade dentro da noção de inteligibilidade não é uma resposta viável, pois não há alternativas que funcionem adequadamente para elucidar o seu funcionamento e que preservem a intencionalidade do fenômeno. Segundo Dennett, quando abandonamos a atribuição de racionalidade, abandonamos também a ideia e a possibilidade de uma interpretação intencional.

A questão que se apresenta neste momento é a respeito do tipo de racionalidade e o modo que ela seria utilizada na noção de inteligibilidade. Seria uma racionalidade procedural, epistêmica ou agencial? Bortolotti (2010) argumenta que se pode fazer sentido de um comportamento mesmo que não satisfaça todos estes critérios. Seriam, talvez, o fechamento dedutivo ou a consistência lógica perfeita? Dennett (1987, p. 94) mesmo concede que são critérios excessivos para o papel em sua teoria. E a adesão a outros princípios da lógica e da probabilidade? É com base nesta discussão que Dennett argumenta que o conceito de racionalidade não é algo que se define facilmente. O mais perto que Dennett concede de uma definição de racionalidade é a seguinte: “Eu quero usar o termo ‘racional’ como um termo de aprovação cognitiva de propósito geral - que requer a manutenção apenas de lealdades condicionais e revisáveis entre a racionalidade, assim considerada, e os métodos propostos (ou até mesmo universalmente aclamados) de progredir, cognitivamente, no mundo.” (Dennett, 1987, p. 97, tradução minha). Seguindo essa diretriz, para ele, o conceito de racionalidade deve ser aquele que é compartilhado – quando compartilhado – sobre o que faz sentido. Para ele não há como fugir do que pré-teoricamente já consideramos como o que faz sentido: ao pensarmos sobre o que mais faz sentido, já estamos nos utilizando dessa forma de pensar.

Dennett (1987, p. 96) propõe que as normas formais advindas de teorias da racionalidade, da lógica e da probabilidade, como teoria dos jogos, são princípios que podem ser utilizados como medidas para a avaliação da racionalidade de argumentos, projetos ou interpretações. Estas normas e teorias, no entanto, não são mais do que indicadores escolhidos, uma forma revisável de se medir a racionalidade. E a racionalidade enquanto conceito, segundo Dennett, tem seu fundamento no que é compartilhado pré-teoricamente a respeito do que faz mais sentido e do que não faz. Desse modo, uma investigação psicológica talvez revelasse que tipo de critério de racionalidade é compartilhado e utilizado implicitamente pelas pessoas, mas esta descoberta empírica revelaria somente mais um dos modos que se utiliza para medir e avaliar a racionalidade.

Logo, para Dennett, ao realizarmos uma interpretação baseada na racionalidade prática como na utilização de inteligibilidade dada por Bortolotti, estamos, inevitavelmente, utilizando um critério de racionalidade que achamos que faz mais sentido. Não precisamos, portanto, deter

o conhecimento das mais corretas maneiras de realizar inferências e estimar probabilidades para atribuímos uma racionalidade ideal, pois ela seria ideal segundo o que *nós* consideramos que é mais correto e faz mais sentido.

Conclusão da análise da noção de inteligibilidade

O meu intuito com esta análise não foi, primariamente, criticar as ideias de Bortolotti, mas utilizar a sua formulação como base para elucidar os pontos mais importantes do aspecto da aplicação da TSI que chamo de nível de previsão.

Em suma, extraí três conclusões ao analisar o modelo de inteligibilidade proposto por Bortolotti. 1) Ao realizarmos interpretações intencionais, às vezes nos utilizamos de crenças pressupostas, implicitamente empregadas na interpretação. 2) Existe uma relação particular entre crenças, desejos e comportamentos que deve ser satisfeita para que a interpretação intencional se complete; esta relação é expressada pela equação intencional. 3) Existe um modelo de racionalidade particular que é utilizado dentro da noção de inteligibilidade; ele se difere das noções formais de lógica e probabilidade por não se utilizar de uma definição explícita, mas ele existe e é necessário, permeando todas as relações genuinamente intencionais do sistema.

Por motivos didáticos resumi os corolários da análise de Bortolotti em três conclusões separadas. Mas, a realidade é que elas não podem ser dissociadas dessa maneira, pois se mesclam e se associam para formar o nível de previsão e explicação do comportamento. Uma forma mais íntegra de expressar as suas características seria a seguinte: para a previsão ou explicação de comportamentos pela postura intencional é necessário encontrar pares de crenças e desejos que satisfaçam a equação intencional. Esses pares de crenças e desejos devem ser encontrados em meio a uma rede holística de crenças, racionalmente estruturada e interligada.

Concluo, portanto, que a noção de inteligibilidade como proposta e explicada por Bortolotti não é precisamente o que eu gostaria de designar como o nível de análise da previsão de comportamentos. Mas a sua análise foi útil para que eu pudesse discriminar mais precisamente o que eu quero designar como essencial neste nível de aplicação. O que fica faltando para que se alcance o devido entendimento da aplicação da postura intencional é a elucidação de *como* encontrar tais pares de crenças e desejos para a previsão do comportamento. Já vimos que crenças implícitas nos ajudam a encontrar estes pares, pois permitem que não precisemos pressupor todas as crenças e realizar todos os passos inferenciais explicitamente.

Aliás, sugiro que elas não só ajudam no processo de descoberta de crenças adequadas para a interpretação como são o principal meio pelo qual isso é feito.

Note-se que uma crença pode ser implícita de dois modos. O primeiro modo em que crenças são implícitas é aquele que pressupõe a completude da racionalidade do sistema. Neste sentido, as crenças implícitas seriam aquelas racionalmente deriváveis de qualquer outra do sistema. Como afirma Schwitzgebel (2018):

Perhaps all that's required to implicitly believe something is that the relevant content be swiftly derivable from something one explicitly believes (Dennett 1978, 1987). Thus, in the planets case, we may say that you believe explicitly that the number of planets is 9 and only implicitly that the number of planets is less than 10, less than 11, etc. (Schwitzgebel, 2018)

O segundo modo em que crenças são implícitas é aquele em que as crenças são atribuídas não porque elas são deriváveis de outras, mas porque elas são atribuídas não-conscientemente. Quando comparando este tipo de crença implícita com a crença implícita no sentido anterior, Schwitzgebel escreve:

Knowledge that is “implicit” in this sense will normally not be implicit in the sense of the previous subsection (if it were swiftly derivable from what one explicitly believes, presumably one could answer the test questions correctly); it's also at least conceptually possible that some such psychologically implicit knowledge may be stored “explicitly” in the sense of the previous subsection. (Schwitzgebel, 2018)

Esta divisão é crucial para os propósitos deste trabalho, pois não é o segundo, mas o primeiro sentido que delimita precisamente o que é função do primeiro e o que é função do segundo nível de aplicação. As crenças *implícitas* no primeiro sentido são aquelas utilizadas pelo nível de previsão e explicação de comportamentos. Estas crenças implícitas fazem parte, ou, melhor, constituem a rede de crenças racionalmente atribuída que sustenta a interpretação intencional.

Já as crenças *explícitas* do primeiro sentido são aquelas em que o nível de atribuição de crenças e desejos se manifesta. Neste nível de aplicação (o primeiro, de atribuição de crenças) são atribuídos os desejos e as crenças iniciais. Isso pode acontecer conscientemente ou não-conscientemente, ou seja, explicitamente ou implicitamente no segundo sentido. Estas crenças são utilizadas como suporte para a derivação de outras crenças racionalmente atribuídas, na aplicação do segundo nível de aplicação.

No exemplo de Bortolotti, as crenças atribuídas de que um chapéu lhe protegeria do vento e que estaria ventando ao sair de casa seriam crenças atribuídas no primeiro nível de atribuição (atribuições aprendidas pelo contexto de Stewart). Mas a crença de que ele, portanto,

poderia se proteger do frio com um chapéu *naquele* momento com *aquele* chapéu, e que o fez vesti-lo antes de sair de casa, é atribuída pelo pressuposto de racionalidade do segundo nível, o de previsão de comportamentos.

4.2.2 O nível de atribuição

Elaboro, agora, o nível de atribuição de crenças e desejos iniciais, o primeiro a ser utilizado em ordem cronológica na aplicação da TSI. Este nível, apesar de não ser a alma da TSI como o nível de previsão de comportamentos, ainda é necessário e tem um papel imprescindível a ser cumprido, pois é nele que são realizadas as atribuições iniciais de crenças e desejos. Determina-se o que o sistema deveria acreditar e desejar de acordo com as suas características e a sua história. Ele é representado pelos passos (1) e (2) na formulação de Dennett da aplicação da postura intencional.

O objetivo da interpretação nesse nível de análise é realizar a atribuição de crenças e desejos ao sistema de maneira que eles possam ser utilizados posteriormente para prever corretamente o comportamento do sistema no nível de previsão. Como realizado na seção anterior, utilizarei as análises efetuadas no segundo capítulo para elucidar o funcionamento deste nível de aplicação.

Começo a investigação observando que, segundo Dennett, qualquer atribuição de crenças e desejos que retorne um resultado positivo em suas previsões pela aplicação da postura intencional será, pelo critério final de avaliação da TSI, uma crença em seu sentido mais completo (Dennett, 1987, p. 15). Pode-se, por conta disso, atribuir *quaisquer* crenças neste nível de atribuição inicial a um sistema desde que faça sentido do comportamento? A resposta de Dennett parece ser que sim, desde que possibilite uma previsão acurada de seus comportamentos. Isso não significa que esta atribuição adequada seja um processo simples. Nos próximos parágrafos lidarei com a tarefa de esclarecer este processo.

Para tanto, retomo parte da discussão do capítulo anterior apontando para o fato que, aparentemente, a atribuição inicial de crenças não é necessária para que a postura intencional possa funcionar. Esta aparência decorre da visão de que a previsão intencional só necessita de um par de crenças e desejos que faça sentido do comportamento. Por que, então, seria *necessário* termos crenças e desejos pré-designados para a interpretação de um sistema? Esta aparência se dissolve rapidamente quando percebe-se que, com toda a gama de possibilidades à disposição de um intérprete, é necessário um modo de encontrar, filtrar e escolher as hipóteses concorrentes para a explicação e previsão do comportamento, ou seja, deve haver um meio de se alcançar a redução da indeterminação da interpretação, que não só é positiva como essencial.

Esta discussão foi realizada no capítulo anterior e originou-se das ideias desenvolvidas por Quine (1960) e adaptadas para a interpretação intencional na crítica de Dub (2015). Segundo Dub, a redução da indeterminação na interpretação é possibilitada por princípios epistêmicos já amplamente utilizados na ciência, como simplicidade, conservadorismo e fecundidade. Para ele não haveria qualquer necessidade de um critério a mais de racionalidade para alcançar esse fim.

Dennett, contudo, reforça que é precisamente com a pressuposição de racionalidade que a TSI irá enfrentar a indeterminação, sendo, justamente, através dela que surge o enorme poder preditivo da TSI. Para afirmar isso, Dennett se utiliza amplamente do argumento evolutivo. Este argumento afirma que o pressuposto de racionalidade utilizado pela TSI é uma aposta segura pois a evolução garante que os organismos selecionados por ela serão, em grande medida, racionais.

O argumento evolutivo

Dennett afirma que a atribuição de racionalidade e a utilização da postura intencional são apostas asseguradas pelo processo histórico de evolução dos organismos alvos de atribuição, pois há uma forte pressão seletiva para a formação do que Dennett (1978a, p. 8) chama de *optimal design* (projeto ótimo). O que esperaríamos de um organismo projetado de maneira ótima? Em um projeto ótimo, segundo Dennett, é seguro pressupor que a maior parte de suas estratégias serão racionais, mas também que a maior parte das crenças de um organismo, especialmente aquelas que dizem respeito ao que é mais importante para o organismo, serão verdadeiras.

Dennett argumenta pela adaptabilidade das crenças verdadeiras em diversos de seus textos (cf. Dennett, 1978a, 1987), mas lida com esta questão diretamente em seu artigo ‘The evolution of Misbelief’ (2009), em co-autoria com Ryan McKay, onde analisa a hipótese da adaptabilidade das crenças falsas. Em primeiro lugar, os autores concedem a seus críticos que a evolução não é uma projetista perfeita de organismos, mas sim sujeita a restrições econômicas, históricas e topográficas. Eles reconhecem, também, que o processo de evolução pode, em alguns momentos, produzir um organismo que possua muitas falsidades em suas crenças sendo que algumas delas poderiam ser adaptativas, mas argumentam que elas seriam adaptativas pelas circunstâncias em que se encontram e que, por si só, não trariam nenhum valor adaptativo, ou seja, elas são, no máximo, *toleradas* pela evolução, e não utilizadas como critério de seleção.

Ao analisar muitas das propostas de circunstâncias potencialmente adaptativas de crenças falsas, Dennett chega à conclusão de que há alguns raros casos de crenças falsas adaptativas, a saber, ilusões positivas. Mesmo assim, os autores afirmam que o truísmo de que informações falsas levam, em geral, a custosos equívocos não foi seriamente enfraquecido. Embora a sobrevivência seja a única moeda que tenha valor para a seleção natural, Dennett e McKay concluem que na maioria das circunstâncias a taxa de troca pela verdade será claramente favorável.

Um dos efeitos disso é que um organismo que apresente a condição em que a maioria de suas crenças são falsas não teria capacidade alguma de sobreviver. É o que Dennett argumenta com a analogia do personagem infantil, Mr. Magoo. Mr. Magoo é um senhor com extraordinária miopia, que se engana a respeito de quase tudo que está a sua volta, saindo da janela achando que é um elevador e caminhando por um pântano infestado de crocodilos achando que são calçadas com algumas poças e sacolas de lixo. A graça é que Mr. Magoo nunca percebe que está prestes a se dar mal porque, por coincidência do destino, as situações sempre se resolvem. “É uma presunção cômica maravilhosa, e acho que isso prova meu ponto. Reconhece-se que o mundo não pode ser assim, nem por mais que alguns minutos. O míope Magoo é impossível por uma razão bastante profunda. Você não pode sobreviver até a idade adulta, a menos que a maioria de suas crenças seja verdadeira.” (Dennett, 1992, tradução minha). É até mesmo difícil conceber um organismo cuja maioria das crenças não seja verdadeira. Dennett argumenta como a ideia inicial de alguém que parece ter muitas de suas crenças falsas, na verdade, nem chega perto da quantidade de crenças verdadeiras que possui.

The idea that most of anyone's beliefs must be true seems obvious to some people. ... A few reflections on peripheral matters should bring it out. Consider Democritus, who had a systematic, all-embracing, but (let us say, for the sake of argument) entirely false physics. He had things all wrong, though his views held together and had a sort of systematic utility. But even if every claim that scholarship permits us to attribute to Democritus (either explicit or implicit in his writings) is false, these represent a vanishingly small fraction of his beliefs, which include both the vast numbers of humdrum standing beliefs he must have had (about which house he lived in, what to look for in a good pair of sandals, and so forth) and also those occasional beliefs that came and went by the millions as his perceptual experience changed. (Dennett, 1987, p. 19)

Elaboro este argumento para esclarecer um ponto de crucial relevância neste momento. O argumento evolutivo estabelece que a maior parte das crenças de um organismo que evoluiu serão verdadeiras *tanto quanto* ele estabelece que as estratégias de um organismo serão, em sua maioria, racionais. Esta conclusão decorre do fato de que o que Dennett (1983) realmente

postula que é a garantia da posse de estratégias racionais e crenças verdadeiras é que este organismo selecionado pela evolução seja *otimamente projetado para os seus fins*.

Se a equação intencional é uma equação de três termos, a crença, o desejo e o comportamento (deixando a capacidade do sistema de lado para simplificar, mas ela poderia ser pensada da mesma maneira), é preciso saber das crenças e desejos de um sistema para descobrir seu comportamento, do seu comportamento e dos seus desejos para descobrir as suas crenças ou das suas crenças e dos seus comportamentos para descobrir os seus desejos. Portanto, é necessária uma drástica redução das possibilidades de resolução para progredirmos na interpretação intencional, e TSI sabe justamente o que utilizar para isto. É com a pressuposição de um projeto ótimo que a TSI obtém, de antemão, boa parte das crenças e dos desejos do sistema para resolver a equação intencional. É por isso que Dennett pode substituir, em seus exemplos de aplicação da TSI, organismos frutos da evolução por artefatos como programas de computador ou termostatos. Porque a pressuposição para os dois é a mesma, que sejam projetados adequadamente para os seus fins e funcione de modo ótimo (Dennett, 1987, p. 73).

Desse modo, saber dos objetivos ou dos fins de um sistema é tão importante quanto saber de suas crenças. Se não soubéssemos *para que* o sistema foi projetado, estaríamos perdidos em nossa interpretação intencional. Para a maioria dos sistemas artificiais temos uma visão privilegiada dos bastidores e conhecemos, mesmo que grosseiramente, o objetivo do sistema, seja ganhar uma partida de xadrez ou manter a temperatura de um ambiente em um valor determinado. Já para organismos que evoluíram fixa-se os seus interesses em sobrevivência, reprodução e qualquer coisa que promova estes dois objetivos, como alimentação, segurança, ausência de dor, sexo, conforto, etc. Realizando-se isso, podemos inferir as crenças de um sapo que pula rapidamente para longe de nós: tendo bons motivos para pressupor que sapos querem fugir de coisas maiores que eles para sobreviver, podemos inferir que ele acredita que somos um tipo de predador (Dennett, 1987, p. 87).

Porém, os desejos de um sistema complexo podem vir a ser tão específicos que estas pressuposições não mais resolvem completamente o problema. No final, uma mistura da atribuição dos desejos e crenças mais básicos e prováveis auxiliarão na busca de pares de crenças e desejos que possam fazer sentido do comportamento e ter a capacidade de pressupor novos. Como esta atribuição e investigação funcionam detalhadamente será abordado na próxima seção.

Conclusão da análise do argumento evolutivo

A partir da análise efetuada, concluo que o argumento evolutivo desemboca em duas principais conclusões pertinentes aos nossos interesses. Primeiramente, o entendimento de que a racionalidade do sistema pode ser vista como uma aposta segura confere um alto grau de confiança de que o *nível de previsão de comportamento* trará resultados positivos em suas previsões. E, por segundo, o argumento evolutivo justifica a expectativa de um *projeto ótimo*, que fornece um método de descoberta para as *atribuições iniciais de crenças e desejos* através de uma pressuposição de posse de crenças verdadeiras e adequadas para os interesses do sistema, permitindo que o segundo nível cumpra a sua função de reduzir a indeterminação da interpretação. Neste método de descoberta de crenças e desejos, pressupõe-se os desejos que são frutos da seleção natural, ou seja, qualquer desejo que tenha gerado uma vantagem adaptativa o suficiente, geralmente relacionados à sua sobrevivência (Dennett, 1987, p.19) e, em seguida, atribuem-se as crenças que seriam necessárias para se alcançar esses objetivos, e, em sua maioria, crenças verdadeiras. A racionalidade pressuposta *para o nível de atribuição de crenças e desejos* não é mais do que a pressuposição de um projeto ótimo do sistema. O que significa que a racionalidade proposta para este nível de aplicação, portanto, não é a racionalidade do sistema em si, mas a racionalidade do “criador” que “projetou” este sistema de uma maneira ótima, isto é, racional.

Proponho que a mistura destes dois alvos de racionalidade é o motivo da confusão teórica de afirmações como a seguinte: “Correspondentemente, a ação de uma pessoa pode parecer irracional quando considerada isoladamente, mas pode ser racional quando é mais amplamente considerada como parte de um preço que vale a pena pelo bom gerenciamento da memória.” (Cherniak 1983, p. 23, tradução minha). Neste caso, a não ser que consideremos que a própria pessoa deseja ter uma gestão adequada de sua memória (o que não seria o caso), proponho que há um equívoco de atribuição de racionalidade à pessoa, quando, na realidade, a racionalidade que faz sentido atribuir-se pela gestão adequada de memória seria para um *segundo* sistema, aquele que teria projetado este primeiro. Em alguns casos seriam engenheiros, mas quando falamos de organismos estamos falando da própria seleção natural enquanto projetista (Dennett, 1995). Assim, em cenários como estes, o que é conferido à pessoa é, justamente, a apresentação de um projeto ótimo que, apesar de possibilitar a atribuição de crenças e desejos iniciais ao sistema, não é a *mesma* racionalidade que possibilitará uma interpretação intencional do sistema pela aplicação do nível de previsão de comportamentos.

Apesar de ser a principal maneira em que a TSI pode determinar a atribuição inicial de crenças e desejos, o argumento evolutivo não é o único modo de se chegar às crenças e desejos que tragam um resultado positivo na sua previsão de comportamentos. Este método pode ser

incrementado e até corrigido por generalizações empíricas aprendidas indutivamente (Dennett, 1987, p. 53). Na verdade, este método, *deve* ser corrigido, pois não só a evolução garante que o sistema será em sua maior parte racional e que ele tenha crenças verdadeiras, ela também garante que o organismo será imperfeitamente racional, no sentido de possuir um projeto imperfeito (Dennett, 1978a, 246). Uma potencial tensão surge entre o método do argumento evolutivo e das generalizações empíricas no momento em que crenças falsas começam a ser atribuídas ao sistema. Para esclarecer funcionamento da interação entre estas orientações que se cruzam, a próxima questão que buscarei responder é, precisamente, qual seria o modo e quais as crenças e desejos que podem ou devem ser atribuídos inicialmente a um sistema.

Como atribuir as crenças e desejos iniciais adequadamente

Dennett (1987) propõe algumas diretrizes para a atribuição de crenças e desejos de modo que os parâmetros estabelecidos pela sua teoria sejam respeitados. Primeiramente, ele propõe atribuir crenças verdadeiras ao que a capacidade perceptiva do sistema lhe proporcionou, tendo em vista a sua história de vida, dentro do campo de análise e nível de resolução dos seus interesses.

Exposure to x, that is, sensory confrontation with x over some suitable period of time, is the normally sufficient condition for knowing (or having true beliefs) about x. In spite of the phrase "know all about," what we come to know, normally, are only all the relevant truths our sensory histories avail us...So one rule for attributing beliefs in the intentional strategy is this: attribute as beliefs all the truths relevant to the system's interests (or desires) that the system's experience to date has made available. (Dennett, 1987, p. 18)

Dennett também aponta para quais desejos deveríamos atribuir inicialmente a um sistema. Seriam aqueles que, levando em conta as suas necessidades biológicas, como sobrevivência e procriação, seriam os mais praticáveis modos de satisfazê-las. Estes seriam desejos como alimentação, segurança, saúde, sexo, riqueza, influência, e qualquer outro que, na visão do sistema, fosse um meio para alcançar estes mesmos desejos (Dennett, 1987, p. 49).

A atribuição de crenças e desejos iniciais também pode ser orientada pelos aspectos gerais e pelas competências que são esperadas do sistema através de uma lista de “especificações”. Estas especificações formariam o *perfil de crença do sistema* que diria que tipo de informação pode ser recebida, usada e transmitida pelo sistema. Neste modo de orientação, utiliza-se o conhecimento do funcionamento interno do sistema e se faz alusão aos modos em que as coisas no mundo devem ser representadas, mas somente em termos utilizáveis

pela postura intencional, ou seja, através de atribuições diferenciadas de crenças e desejos (Dennett, 1987, p. 255). Assim, por exemplo, o conhecimento dos aparatos perceptivos de um sapo e dos seus mecanismos de discriminação de cor e movimento motivaria uma interpretação diferenciada pois seria a diferença de atribuirmos uma crença de que “algo preto comestível se move nesta direção”, de “uma mosca varejeira está voando na minha frente”. Do mesmo modo, pode-se montar um perfil de crenças de um sistema baseado na classificação de graus que esse sistema teria com relação à posse do que filósofos e psicólogos chamam de *Teoria da Mente*. Por exemplo, Dennett (1988) discute qual seria o perfil de crenças de primatas *Cercopithecidae* (ou Macacos do Velho Mundo) quando discute se eles seriam sistemas intencionais de primeiro, segundo ou terceiro grau, como exposto no capítulo anterior.

Estes são alguns dos princípios que Dennett estabelece para a atribuição de crenças. Dennett busca regras generativas que possam assistir na atribuição de crenças em diferentes situações, por mais diversas que sejam. No entanto, o conhecimento a respeito de que tipo de crenças atribuir ou não a um sistema pode ser incrementado ou corrigido de acordo com as evidências empíricas a respeito do que mais funciona.

Por exemplo, Dennett afirma que a exposição para as coisas é só *normalmente* suficiente para o conhecimento delas. Em outras circunstâncias anormais, mesmo tendo em vista esta exposição, o nosso limite para aceitar ignorância pode ser consideravelmente grande. Para manter a integralidade do sistema, Dennett aponta alguns parâmetros para atribuir crenças falsas.

But the attribution of false belief, any false belief, requires a special genealogy, which will be seen to consist in the main in true beliefs. ... The falsehood has to start somewhere; the seed may be sown in hallucination, illusion, a normal variety of simple misperception, memory deterioration, or deliberate fraud, for instance, but the false beliefs that are reaped grow in a culture medium of true beliefs. (Dennett, 1987, p. 18)

Neste sentido, não são regras fixas e imutáveis. O conhecimento pode variar de regras generalizáveis a quase qualquer situação, como as crenças perceptivas dos sistemas exposto por Dennett, até regras mais pontuais para situações particulares, como a crença das crianças pequenas de que Papai Noel existe (Dennett, 1987, p. 54). Em outra passagem, Dennett escreve como teorias psicológicas sobre o funcionamento do processamento de informações podem determinar o modo como atribuímos crenças e desejos, mas desde que se respeite a racionalidade esperada de não haver nenhuma contradição na rede de crenças do sistema (Dennett, 1987, p. 88).

Dennett escreve sobre como esse tipo de conhecimento também pode ser aplicado para ajudar na aplicação da TSI para casos complicados de irracionalidade:

Having learned about “cognitive dissonance,” Stich is now prepared to find both in himself and in others the resolution of cognitive dissonance in the favoring of a self-justifying belief over a less comfortable belief better supported by the evidence. This is a fine example of the sort of empirical discovery that can be used to tune the intentional stance, by suggesting hypotheses to be tested by the attributer. (Dennett, 1987, p. 99)

Tendo em mente todas as considerações realizadas sobre o nível de aplicação de atribuição, agora podemos entender mais claramente o projeto de Dennett para a atribuição de crenças e desejos neste nível de aplicação. Para finalizar a explicação deste nível, eis uma passagem do autor sobre a aplicação da postura intencional para Macacos do Velho Mundo que resume a sua proposta.

My proposal, in simplest terms, was this. First, observe their behavior for a while and make a tentative catalogue of their needs—their immediate biological needs as well as their derivative, informational needs—what they need to know about the world they live in. Then adopt what I call the intentional stance: treat the monkeys as if they were — as they may well turn out to be — rational agents with the “right” beliefs and desires. (Dennett, 1988, 207)

Resumo dos níveis de aplicação

Com a explicação dos dois níveis tendo sido elaborada, pode-se delinear explicitamente o funcionamento da aplicação da postura intencional para qualquer sistema. Há, primeiramente, a aplicação do nível de atribuição, em que há atribuições iniciais de crenças e desejos. Estes têm o objetivo de reduzir a indeterminação da interpretação e possibilitar a descoberta de pares de crenças e desejos que vão formar os elementos para a previsão ou explicação do comportamento. Generalizações empíricas e conhecimento indutivo podem (e devem) ser utilizados para o ajuste fino das atribuições, mas o que vai nortear a busca de crenças e desejos a serem atribuídos inicialmente e realizar o trabalho pesado é a pressuposição de um projeto ótimo, justificado pelo argumento evolutivo e pela confiança na seleção natural (Dennett, 1971, p. 92).

Em segundo lugar, emprega-se o segundo nível de aplicação, o nível de previsão, em que se explica o comportamento sob análise ou se prevê o comportamento futuro do sistema. Busca-se, para este fim, pares de crenças e desejos que sejam racionalmente derivados das

crenças e desejos inicialmente atribuídos que possam prever ou explicar um comportamento através da equação intencional.

4.2.3 Considerações importantes

Nesta seção, exponho considerações fundamentais para o correto entendimento desta intrincada tarefa da aplicação da TSI para casos de irracionalidade e controversa atribuição de crenças. Estas considerações referem-se à relação causal das crenças, desejos e outros estados mentais. Explorarei a interação entre os elementos da postura intencional e sentimentos e emoções para entender um pouco melhor como a TSI lida com as causas e os efeitos das crenças quando não estão em uma interação intencional racionalista.

Crenças e desejos como causas não intencionais

A relação intencional existente entre os elementos constitutivos da TSI já foi explorada no presente trabalho. Contudo, ainda há o que se investigar a respeito das origens das crenças e dos desejos e também de seus destinos, ou seja, as suas causas e os seus efeitos. Assim, o que eu explorarei na presente seção é a maneira que a TSI lida com a possibilidade de que crenças e desejos possam ser gerados por fenômenos não intencionais e com a possibilidade de que eles também sejam produtos que não sejam pautados pela intencionalidade.

Recapitulando: segundo a teoria de Dennett, utilizamo-nos de três posturas diferentes para entender os objetos no mundo, a saber, a postura física, a postura de projeto e a postura intencional. De maneira similar, podemos entender a produção e os efeitos das crenças pelas mesmas posturas. Já vimos na seção 2.4.2 que nem todas as explicações que contêm termos intencionais são, *ipso facto*, explicações intencionais (Dennett, 1978a, p. 236). Por meio da postura física ou de projeto, podemos utilizar uma crença como a causa de outros fenômenos no mundo, como um ataque cardíaco, atos falhos ou o ruborizar-se (Dennett, 1987, p. 57). Mas como fazer sentido de que o conceito de crença proposto por Dennett, que é um construto lógico com o propósito de servir no cálculo do comportamento intencional de um sistema, pode causar fisicamente outros fenômenos no mundo?

A resposta para Dennett é que assim como não é necessário ser um realista identitário sobre centros de gravidade para citá-los como causas (que podem fazer com que um barco seja instável, por exemplo) e assim como não é necessário ser um realista sobre outros fenômenos difusos e de alto nível como as pressões seletivas dentro da teoria evolucionista, ou ainda como

a inflação em economia, também não é necessário ser um realista identitário sobre crenças para citá-las como causas de faces coradas, por exemplo (Dennett, 2000, p. 358).

Relação da postura intencional com emoções e sentimentos

Em seus textos sobre a TSI, Dennett quase não escreve sobre emoções e sentimentos, mas pode-se inferir das passagens analisadas que Dennett não teria problema com a noção ordinária de que crenças causam emoções. Do mesmo modo que posso enrubecer se acreditar que estão olhando para mim e me julgando, posso também, pelos mesmos motivos, me sentir envergonhado. Na verdade, Dennett reconhece que emoções têm um papel importante na busca das crenças de um sistema, pois podem sugerir quais crenças um sistema possui. Suspeitaríamos de um cirurgião que diz que tudo vai dar certo em nossa cirurgia enquanto apresenta claros sinais de medo e ansiedade como voz trêmula, suor ou palidez. Isso acontece porque reconhecemos que as emoções podem ser preditores confiáveis de crenças ou desejos. Dennett pensa em se utilizar dessa estratégia para entender quais crenças seriam mais prováveis um chimpanzé possuir em um experimento, já que eles não podem relatar as suas crenças. Para Dennett, evidência de que o chimpanzé estivesse perplexo seria um bom sinal de que ele tem crenças que contradizem com o que ele está percebendo no momento (Dennett, 1987, 255). Uma explicação mais completa da estratégia que estou destacando está na seguinte passagem:

Very young children watching a Punch and Judy show squeal in anticipatory delight as Punch prepares to throw the box over the cliff. Why? Because they know Punch thinks Judy is still in the box. They know better; they saw Judy escape while Punch's back was turned. *We take the children's excitement as overwhelmingly good evidence that they understand the situation*—they understand that Punch is acting on a mistaken belief (although they are not sophisticated enough to put it that way). Would chimpanzees exhibit similar excitement if presented with a similar bit of play acting (in a drama that spoke directly to their "interests")? I do not know, and think it would be worth finding out for if they didn't react, the hypothesis that they impute beliefs and desires to other would be dealt a severe blow, even if all the P&W [Premack & Woodruff, 1978] tests turn out positively, just because it can be made so obvious--obvious enough for four-year-old children—that Punch believes (falsely) that Judy is in the box. (Dennett, 1978b, p. 570)

Portanto, indicações de emoções podem ser boas aliadas na busca de quais crenças e desejos podemos atribuir a um sistema em situações complexas de irracionalidade. É importante, no entanto, delimitar o conceito de emoções e sentimentos de um modo que se encaixe na teoria geral de Dennett sobre os estados mentais.

A tarefa proposta aqui não é de elaborar uma análise extensa a respeito da natureza desses estados mentais, mas somente entender a sua relação com a previsão de comportamentos a partir da TSI, bem como sua influência sobre esta. Portanto, quando discorro a respeito dos efeitos e características das emoções e sentimentos, o faço apenas para ressaltar o que é possível na interação com crenças e desejos dentro da TSI, não para avançar em alguma discussão a respeito de teorias psicológicas a respeito das emoções.

Apesar de Dennett não ter escrito especificamente sobre emoções, podemos entender um pouco de como a TSI lida com emoções e sentimentos extraindo alguns conceitos gerais pela análise que ele realizou sobre o conceito de dor. Para Dennett, há muita confusão sobre o conceito de dor que é intuitivamente utilizado pela PSC, pois busca conciliar os efeitos que a dor teria no organismo e nos estados intencionais de uma pessoa com os estados fenomenológicos que acompanham o sentimento de dor. Assim, nem tudo o que a PSC entende sobre o fenômeno da dor será coerente e adequado para ser incorporado pela TSI.

Apesar da confusão teórica que pode haver em torno do conceito de dor, para Dennett, há alguns efeitos claros extraídos do conceito de dor que dizem respeito aos centros superiores da mente, como crenças, desejos e o raciocínio. Dennett (1978c) observa que quando temos dor, também acreditamos que temos essa dor e podemos lembrar dela por algum tempo. As dores interferem, também, na nossa habilidade de concentração, de resolver problemas e de pensar claramente. Mas, de um modo mais importante, as dores são aversivas e são modificadores confiáveis de comportamentos, pois alteram as nossas intenções e desejos quando estão presentes. Assim, podemos entender melhor o que Dennett pensa a respeito da interação de emoções e sentimentos com crenças e desejos. As dores modificam os nossos desejos, pois são aversivas, fazendo com que busquemos nos afastar delas e que evitemos o que as promove. Além disso, as dores também causam crenças a respeito de sua existência agora e também sobre a sua existência no passado. Desse modo, há espaço dentro da teoria de Dennett sobre crenças para que emoções e sentimentos modifiquem nossas crenças e nossos desejos.

Um ponto importante que precisa ser ressaltado é que o conhecimento que se tem sobre o efeito da dor vem de um conhecimento indutivo e empírico sobre o que se conhece sobre o mundo. Desse modo, emoções e sentimentos não são categorias conceitualmente necessárias para a análise dos efeitos da vida mental de uma pessoa sobre o seu comportamento. Pois, assim como afirma Dennett:

Pains are also goal modifiers , but they might not be. That is, we can imagine a person who says he is in pain, locates the pain consistently, is in fact being beaten , writhes, cries, trembles, but is immune to torture . Is this really imaginable? Of course it is.

Perhaps that is what masochists are. Or perhaps they have, as it were, a sign reversed going to the goal box, so they seek out pain instead of avoiding it, at least in certain circumstances. (Dennett, 1978a, p. 220)

A direção de influência entre crenças e desejos e emoções é que crenças causam emoções e que emoções influenciam desejos: por exemplo, ficamos com raiva quando nos fazem uma injustiça, o que gera o desejo de revidar. Que este seja o caso, no entanto, é uma questão de aprendizado empírico. Como destacado, é plausível que haverá casos em que emoções causem crenças ou que desejos causem emoções.

Esta digressão sobre a relação entre crenças, desejos e emoções foi feita para destacar que não há regras pré-estabelecidas, dentro da TSI, para o que pode ser considerado como gerador de crenças e desejos e o que estes podem causar. O que eu quero evidenciar são dois pontos sobre a busca de crenças e desejos para o nível de atribuição. Primeiro, esta busca pode ser auxiliada avaliando os efeitos que crenças e desejos podem ter no sistema, como emoções ou o enrubescer de faces, que não sejam unicamente os comportamentos intencionais. Segundo, podem ser utilizadas ambas as posturas de projeto e física para esta busca de atribuição inicial, seja através de emoções que gerem desejos particulares ou a mera exposição a um evento que gere uma crença sobre ele.

4.2.4 Considerações para casos de irracionalidade

Para elaborar de maneira mais clara como a postura intencional pode ser aplicada para casos de irracionalidade, retomo um exemplo de Dennett, mencionado no capítulo anterior, sobre o efeito da distração na nossa capacidade de entender intencionalmente o nosso comportamento. Neste exemplo, Dennett se esquece do jogo de tênis que havia marcado para o horário do almoço com seu amigo e vai almoçar, em vez de ir para a quadra. Quando confrontado por seu parceiro de jogo, Dennett pede desculpas e responde que se esqueceu completamente do combinado. Como entender o que aconteceu pela postura intencional? Creio termos agora as ferramentas necessárias para entender como podemos utilizar a TSI para lidar com tal falha em sua rede de crenças. Fazemos isso utilizando o nosso conhecimento de como a memória e a atenção funcionam para interpretar o comportamento de Dennett.

Usemos uma compreensão rudimentar de como a atenção funciona: atenção é o mecanismo que filtra e nos indica quais informações devemos e estamos processando dentre toda a gama de informações disponíveis para o organismo. Utilizando tal modelo (meramente

ilustrativo), podemos entender que o que aconteceu com Dennett é que ele deixou de processar a informação de que teria um compromisso no almoço ao planejar as suas atividades na hora do almoço. Mas como isso ajudaria na nossa compreensão do motivo de ele ter ido almoçar? Esta compreensão só ajudará se ela possibilitar uma modificação em nossa atribuição de crenças à Dennett. Esta alteração tem um caminho “artificial” para a sua rede de crenças, mas é precisamente desse modo que podemos, agora, atribuir-lhe as crenças que façam sentido do seu comportamento. De manhã, quando Dennett combinou o jogo com o seu amigo, atribuímos a ele a crença de que ele tem um compromisso com o seu amigo no horário do almoço. Quando Dennett deixa de prestar atenção no seu jogo por algum motivo, podemos atribuir-lhe a ausência desta crença e, portanto, ele agiria como se não acreditasse que tem o compromisso. Quando a sua atenção se volta para o compromisso, com a conversa com o seu amigo, ele volta a processar a informação relevante. Isso explica o seu comportamento porque, neste momento, atribuímos novamente esta crença a ele, fazendo com que ele conclua que esqueceu do combinado e peça desculpas. Assim, compreender adequadamente a interação entre explicações pela postura intencional e explicações por outras posturas é fundamental para a correta interpretação de sistemas que demonstrem alguma irracionalidade.

In practice we predict lapses at the intentional level ("You watch! He'll forget all about your knight after you move the queen") on the basis of loose-jointed inductive hypotheses about individual or widespread human frailties. These hypotheses are expressed in intentional terms, but if they were given rigorous support, they would in the process be recast as predictions from the design or physical stance. (Dennett, 1978a, p. 246)

Que a sua crença estivesse, de algum modo, em sua mente é uma interpretação possível, mas só faz sentido como uma implementação da postura de projeto para competências da postura intencional. Dentro de uma implementação da postura intencional, crenças contraditórias ou erros de inferência impediriam o seu funcionamento adequado.

4.2.5 A aplicação da TSI como um quebra-cabeça

Para ajudar na visualização de como funciona o modelo proposto em dois níveis da aplicação da postura intencional, proponho o emprego de um exercício de imaginação. Sugiro realizar uma analogia entre a aplicação da postura intencional e a montagem de um quebra-cabeça.

Imagine que o sucesso de uma interpretação intencional pela TSI dependa da formação de uma imagem, uma estrutura esquemática, que conecte a capacidade de ação de um sistema

com a motivação para esta ação. Para se formar esta imagem, é preciso montá-la juntando-se peças que se encaixam de modo particular, porém não exclusivo, assim como em um quebra-cabeça. Estas peças seriam as crenças, que só podem se encaixar umas nas outras pelo uso da racionalidade. O conjunto de peças encaixadas é o que chamamos de rede de crenças. Algumas peças são especiais porque estão diretamente ligadas aos desejos do sistema. Outras peças especiais são aquelas ligadas a uma capacidade de ação do sistema. Ao ligar uma peça de um desses dois conjuntos especiais a uma peça do outro podemos montar uma imagem que revele um comportamento.

Para que se possa começar a montar uma imagem com o quebra-cabeça, algumas peças são colocadas de antemão, ou seja, já há regras ou indicações que apontem onde se encontram, ou podem se encontrar, as primeiras peças para facilitar na montagem do quebra-cabeça. As peças são indicadas em uma fase inicial em que também é indicado o formato do espaço a ser preenchido para a imagem. As peças que darão formato para a imagem serão as *peças de canto*, que às vezes são colocadas nesta fase inicial.

Proponho este exercício de imaginação porque ele é útil em ilustrar alguns pontos que são importantes de serem ressaltados para a presente formulação:

1. O encaixe de uma peça às outras só pode acontecer com o uso da racionalidade. Sem ela as peças estariam soltas e desconectadas umas das outras.
2. As peças desse quebra-cabeça intencional são crenças, e o conjunto das peças unidas é a rede de crenças.
3. A combinação de peças dependerá das peças disponíveis que se encaixem, podendo existir mais de uma combinação possível. Qual montará a melhor ou mais nítida imagem é uma questão a ser julgada.
4. A conexão entre as peças serve para montar uma imagem. Quanto maior e mais distinta é a imagem, mais entendemos a respeito do que ela descreve. Isto é, esta junção das peças se torna a rede de crenças e permite entender um perfil disposicional de comportamentos mais completo para o sistema.
5. É preciso haver uma conexão de peças entre aquela que está conectada ao comportamento e aquela que está conectada ao desejo, ou seja, para que se possa explicar um comportamento intencionalmente é preciso respeitar a equação intencional: $\text{Capacidade (A)} + \text{Crença (A} \rightarrow \text{B)} + \text{Desejo (B)} = \text{Comportamento (A)}$.
6. A tarefa de se montar um quebra-cabeça é possibilitada em grande medida pela redução das possíveis combinações de peças que é oferecida pelo uso das peças iniciais colocadas na fase inicial do quebra-cabeça e do formato indicado. Esta fase inicial seria

o nível de atribuição inicial. As peças encaixadas neste momento possibilitam uma drástica redução nas possibilidades combinatórias entre as peças para que o quebra-cabeça possa ser resolvido em tempo hábil. As crenças que atribuímos a um sistema funcionam de maneira similar. Elas estão ali para a redução da indeterminação para que o problema possa começar a ser resolvido, possibilitando a escolha mais acertada entre diferentes tipos de combinações.

7. O uso das “peças de canto” no quebra-cabeça seriam as crenças que mais facilmente são atribuídas ao sistema, montando também o formato que esperamos ter de imagem, ou seja, o formato do seu *perfil de crenças do sistema*. Por exemplo, se sabemos que crianças menores de quatro anos ou certos macacos não tem a capacidade de ter crenças de segunda ou terceira ordem, então limitamos as nossas atribuições de crenças, ou seja, limitamos o tamanho ou o formato da imagem que queremos formar.
8. Ao mesmo tempo, essas peças iniciais não podem ser colocadas de modo que elas não se encaixem entre si, ou que as peças que se encaixem nelas também sejam conflitantes (pelo menos para utilizarmos um conjunto particular de peças). Se estas peças não se encaixarem, teremos não uma, mas duas ou mais imagens separadas, o que resulta em duas ou mais interpretações intencionais. Neste sentido, deve haver uma concordância entre os formatos produzidos pelas peças inicialmente colocadas e as peças que se encaixam nelas. Do mesmo jeito, as crenças atribuídas inicialmente a um sistema não podem ser conflitantes e será problemático se elas produzirem outras crenças que conflitem entre si.
9. Desenhos incompletos podem ser utilizados para a explicação de alguns comportamentos por meio de um perfil disposicional incompleto. Esta situação gera uma redução na amplitude do que pode ser explicado e previsto e, também, gera instabilidade, pois o pequeno conjunto de peças estaria demasiadamente suscetíveis a não se encaixar com outras peças que se sabe que pertencem ao quebra-cabeça.

4.2.6 Retomada das críticas à TSI

Criei a divisão da aplicação da postura intencional em dois níveis para viabilizar a formação de um método capaz de avaliar a viabilidade da aplicação da postura, mas, como vimos, esta também é uma ferramenta que possibilita o esclarecimento de algumas questões teóricas sobre a TSI. Continuo este esclarecimento revisando as principais críticas direcionadas à TSI.

O argumento do duplo padrão de racionalidade

Comecei a investigação da possibilidade da TSI em lidar com casos de irracionalidade a partir da análise da crítica de Stich e a separação que ele identifica dentro da teoria de Dennett: a linha rígida e a linha suave. Recapitulando: para Stich, a linha rígida insiste que a pressuposição de racionalidade ideal da TSI deve ser encontrada dentro da prática popular de onde a TSI se deriva, enquanto que a linha suave propõe alterações com a noção idealizada de um sistema intencional para trazer mais em harmonia com a prática popular, que não invocaria considerações de racionalidade. A resposta de Dennett a essa diferenciação é que a TSI, na verdade, trabalha com a linha *flexível*, a qual afirma que as pressuposições de racionalidade devem aparecer nas práticas populares (i.e., PSC) mas que a racionalidade não é aquilo que aparenta ser para alguns teóricos, precisando de alguns ajustes.

Embora Dennett explique do seu modo como gostaria que a linha flexível funcionasse, podemos utilizar a ferramenta aqui desenvolvida para facilitar a argumentação de alguns pontos. Entende-se, agora, que a rigidez da linha flexível, isto é, a sua parte que afirma que as pressuposições de racionalidade devem perdurar, se encontra na aplicação do nível de previsão de comportamento, em que, realmente, há uma intolerância quanto à manifestação de qualquer irracionalidade. A flexibilidade da linha, a sua parte suave, se encontra dentro do nível de aplicação da atribuição, em que a racionalidade ideal não é aplicada do mesmo modo. Ela é aplicada ao “projetista” do sistema, isto é, à seleção natural, como um argumento que justifica a pressuposição de um projeto ótimo como principal meio de reduzir a indeterminação da interpretação para a adequada atribuição de crenças e desejos. Neste sentido, o emprego de generalizações que se valem da irracionalidade para auxiliar na atribuição inicial não é só possível como apreciado.

O argumento da dissimilaridade com a PSC

Já vimos que o argumento da dissimilaridade da TSI com a PSC não se segue porque os casos levantados contra a capacidade explicativa ou preditiva da TSI, na realidade, também se aplicam para a PSC. Há somente uma aparência de sucesso na explicação da PSC porque conseguimos “remendar” o rasgo feito na explicação intencional com uma explicação a partir da postura de projeto ou física. Agora entendemos claramente como esse remendo é feito:

devolvendo-se a racionalidade ao sistema a partir de atribuições diferenciadas de crenças e desejos cuja explicação é empreendida via a postura física ou de projeto.

É importante ressaltar também que esse rasgo remendado não trará um impacto expressivo de modo a impossibilitar a nossa interpretação, desde que se pressuponha que todo o resto do tecido da rede de crenças do sistema ainda se mantenha íntegro (o que não aconteceria sem um pressuposto de um fundo de racionalidade, ou seja, de um pressuposto de projeto ótimo).

O argumento do critério arbitrário de redução da indeterminação

Para Dub (2015), o critério de redução da indeterminação pelo uso da racionalidade é arbitrário e indevido. Nenhuma outra ciência se utiliza de um critério de racionalidade para esta função e muito menos de racionalidade ideal. A defesa de Dennett para esta crítica foi observar que este é o método que a TSI utiliza, pois explora a característica que organismos que evoluíram pela seleção natural têm, em sua grande maioria, crenças verdadeiras e estratégias racionais de formação de crenças.

Concluo que uma análise mais apurada da questão revelou que a pressuposição de racionalidade para a redução da indeterminação realmente não é necessária para que esse fim seja alcançado, sendo que outros métodos da ciência empírica já se ocupam desta tarefa. Mas de modo nenhum ele seria uma escolha arbitrária, pois Dennett expõe convincentes motivos para a sua escolha da pressuposição de racionalidade e o impacto que ela tem, viabilizando a maioria de nossas interpretações intencionais por levantar a carga mais pesada da tarefa de atribuição de crenças e desejos, deixando que métodos empíricos possam refinar as atribuições por uma espécie de ajuste fino.

O argumento da definição arbitrária

Outro argumento de Dub contra o critério de racionalidade da TSI é o argumento da definição arbitrária. Para ele, não há um motivo especial ou particular para usar um critério de racionalidade para a atribuição de crenças. Desse modo, debater se um estado mental que apresente irracionalidade é uma crença ou não, não seria muito diferente de discutir se um *wrap* é ou não é um sanduíche. Isto é, o debate seria puramente terminológico.

Mesmo Dennett admite que os critérios para a utilização da postura intencional parecem ser perigosamente circulares (1987, p. 50). A partir da análise realizada, fica evidente como a

racionalidade é inerente a qualquer interpretação intencional pelo nível de previsão. Qualquer interpretação que possa prever o comportamento ou estados de um sistema sem a pressuposição de uma racionalidade, será uma interpretação pela postura física ou de projeto. Assim, parece difícil encontrar algo que pudesse falsear a sua teoria.

O que salva Dennett desta crítica é a real possibilidade de utilização da TSI para produzir hipóteses independentes e testáveis em comparação com outros métodos. Isso foi trabalhado em ‘Intentional systems in cognitive ethology: The “Panglossian paradigm” defended’ (Dennett, 1983) e é, justamente, uma das metas do presente trabalho para as próximas seções.

4.3 Método de aplicação da postura intencional

Com a análise formulada até então, podemos desenvolver os seguintes passos para a aplicação da postura intencional. Primeiramente, realizamos a aplicação do primeiro nível de aplicação, o nível de atribuição de crenças e desejos iniciais. Para atribuir crenças e desejos ao sistema, explora-se o que se sabe sobre ele com relação ao seu funcionamento, ao seu lugar no mundo e à sua história de vida (Dennett, 1987, p. 17, p. 49). Depois, atribuem-se as crenças e os desejos iniciais ao sistema, começando com a pressuposição de um projeto ótimo para o seu lugar e o propósito do sistema no mundo.

Atribuem-se desejos ao sistema com base nas suas necessidades biológicas mais básicas e os meios mais praticáveis de satisfazê-las. Atribuem-se, também, as crenças que são verdadeiras e relevantes para a sua vida, levando em consideração o que foi possibilitado pela sua história de vida e pelo seu sistema perceptivo. Em seguida, esta atribuição inicial de crenças e desejos é incrementada e corrigida por conhecimentos particulares aprendidos empiricamente das características do sistema e do seu funcionamento como um todo, seja com a adição de crenças ou desejos específicos em circunstâncias especiais ou com a especificação de um *perfil geral* de crenças para o sistema.

Ressalto aqui que a presente formulação de um método para a aplicação da postura intencional tem o objetivo de salientar as características da teoria dos sistemas intencionais para a análise de sua viabilidade e não de contemplar todos os fenômenos psicológicos presentes na interpretação real da PSC. Trazer demasiada complexidade para a discussão, no momento, resultaria em maior confusão conceitual do que ajudaria no esclarecimento dos fundamentos teóricos da aplicação da postura intencional. Portanto, utilizo esta ressalva para justificar

utilização de apenas crenças e desejos pertinentes para a interpretação que são cabíveis em uma análise pontual, mesmo entendendo que seria necessário um aparato muito mais sofisticado para destrinchar, mesmo que esquematicamente, o funcionamento da interpretação humana. Posto isto, temos o seguinte esquema para o nível de atribuição:

- 1) Especificação do contexto
 - a. Constatação da sua história e do seu lugar no mundo
- 2) Nível de atribuição de crenças e desejos iniciais
 - a. Desejos
 1. Os meios mais praticáveis de satisfazer as suas necessidades básicas
 2. Atribuição particulares aprendidas de desejos
 - b. Crenças
 1. Atribuição de crenças verdadeiras relevantes para o sistema que foram disponibilizadas pelo seu aparato perceptivo
 2. Atribuição particulares aprendidas de crenças

Partimos, então, para o nível de previsão de comportamentos. Neste nível, é importante recordar que a previsão de comportamentos, apesar de enfatizada, não é o único objetivo para o qual a TSI pode ser utilizada, pois ela também é aplicada para a *explicação* de comportamentos. Apesar de a previsão ser o maior enfoque de Dennett em seus textos (cf Dennett, 1987, p. 49), há uma clara equivalência em seus textos entre as duas finalidades da TSI (cf. Dennett, 1987, p. 259; 1994, p. 24). Apesar desta equivalência, elas demandam e fornecem informações diferentes para o sucesso da interpretação, portanto, diferenciarei levemente os dois modos dentro da aplicação, disponibilizando um passo para cada.

Para a explicação do comportamento, seguimos o direcionamento de Dennett: “Mas quais dessas [proposições] é de fato acreditada será determinado pela observação do que [o sujeito] diz e faz, e então perguntando-se: quais crenças e desejos tornariam aqueles atos racionais?” (Dennett, 1987, p.100). Ou seja, há a necessidade de explicar o comportamento de modo a encontrar os termos que satisfaçam a equação intencional: “Crença (A → B) + Desejo (B) + Capacidade (A) = Comportamento (A)”. Neste ponto, já estamos aplicando o nível de previsão de comportamentos, o que significa que há a necessidade da aplicação de racionalidade para a descoberta de novas crenças, visando a formação de uma rede de crenças internamente racional e coerente.

O objetivo é encontrar o par crença–desejo que explique o comportamento sob análise. Para isso, precisamos de uma ligação até as crenças e desejos iniciais. Há, portanto, três tipos de peças para montarmos este quebra-cabeça: as crenças e desejos iniciais, atribuídos no primeiro nível, as crenças e desejos intencionais, que explicam o comportamento, e as crenças racionais, que são inferidas das outras crenças. Utilizarei estes rótulos para me referir às crenças e aos desejos relevantes utilizados na aplicação, levando em conta que as crenças e desejos intencionais podem ser tanto aqueles inicialmente atribuídos quanto aqueles derivados destes. Mais do que um par de crença e desejo pode ser encontrado, o que resulta na formação de várias hipóteses que competem pela melhor explicação. Temos, então:

3) Especificação do comportamento a ser analisado

4) Nível de previsão

- a. Explicação – Busca do par de crença e desejo intencional que 1) explique o comportamento (através da equação intencional); e 2) seja compatível com o perfil de crenças e desejos e com as crenças e desejos iniciais por uma ligação racionalmente estabelecida.

- Comportamento 1:
 1. Crença inicial 1:
 2. Crença inicial 2:
 3. ...
 4. Crença racional 1:
 5. Crença racional 2:
 6. ...
 7. Crença intencional:
 8. Desejo intencional:
- Comportamento 2:
 1. ...

Agora, passamos da explicação para a previsão de comportamentos. Em si, elas são processos separados e ocorrem, também, separadamente. Porém, levando em consideração a forma e a frequência que elas são utilizadas no cotidiano, utilizar um método que agregue as duas formas e permita que elas sejam utilizadas em conjunto parece ser o mais apropriado. Posiciono, assim,

a explicação de comportamentos seguida da previsão, de forma que as crenças e desejos atribuídos na explicação também possam ser utilizadas na previsão.

Para a previsão de comportamentos dentro da aplicação da postura intencional, buscam-se pares de crenças e desejos que possam satisfazer os termos da equação intencional dentro do que foi atribuído de crenças e desejos em conjunto com a rede de crenças racionalmente inferida: “Crença ($A \rightarrow B$) + Desejo (B) + Capacidade (A) = Comportamento (A)” e, assim, obter o comportamento (A). Diversos comportamentos podem ser compatíveis com a rede de crenças do sistema, logo, como na explicação dos comportamentos, várias hipóteses são levantadas e disputam pelo espaço no perfil disposicional de comportamentos do sistema.

- b.** Previsão – Busca de um comportamento que resulte de um par de crença e desejo intencional que: 1) preveja algum comportamento (através da equação intencional); e 2) seja compatível com o perfil de crenças e desejos e com as crenças e desejos iniciais por uma ligação racionalmente estabelecida.

- Hipótese 1:
 1. Crença inicial 1:
 2. Crença inicial 2:
 3. ...
 4. Crença racional 1:
 5. Crença racional 2:
 6. ...
 7. Crença intencional:
 8. Desejo intencional:
- Hipótese 2:
 1. ...

Qualquer par de crenças e desejos que possa satisfazer as condições acima mencionadas entra na corrida para disputar o lugar de melhor interpretação para a situação sob análise. A decisão acontece à medida em que hipóteses de comportamentos são confirmadas ou descartadas por observações posteriores do sistema. Pode-se rever as hipóteses sobre as previsões do comportamento (passo 4) quantas vezes forem necessárias. Caso todas as hipóteses levantadas

sejam falseadas, então pode-se revisitar a explicação dada e buscar por novos pares de crenças e desejos.

Não só podem ser feitas novas hipóteses para a previsão e explicação do comportamento como há também a revisão das atribuições de crenças e desejos iniciais (passo 2). Em resposta a uma crítica Stephen Webb à TSI, Dennett ressalta:

It overlooks a key feature of the intentional stance: one is allowed to revise one's attributions in the light of falsified predictions. ... One patiently gathers behavioral evidence (largely but not entirely verbal behavioral evidence), hypothesizing interpretations and refining one's attributions until, in the limit, they account for ("predict," make sense of) all the behavior (Dennett, 1994a, p. 24)

Alvos especialmente predispostos à revisão são as crenças que são dependentes das condições de posse de informação do sistema, como as crenças perceptivas e as advindas da memória do sistema.

The presumption of rationality is so strongly entrenched in our inference habits that when our predictions prove false, we at first cast about for adjustments in the information-possession conditions (he must not have heard, he must not know English, he must not have seen x, been aware that y, etc.) or goal weightings, before questioning the rationality of the system as a whole. (Dennett, 1978a, pp. 9-10)

Há também a consideração que cada reformulação e interpretação funciona como uma hipótese provisória, em competição pela melhor alternativa interpretativa.

These particular fall-back positions are themselves subject to the usual tests on belief attribution, so merely finding a fall-back position is not confirming it. If it is disconfirmed, the search goes on for another saving interpretation. If there is no saving interpretation — if the person in question is irrational — no interpretation at all will be settled on. (Dennett, 1987, p. 87)

Portanto, quando todas as nossas previsões são frustradas e as novas hipóteses se mostram insatisfatórias, abandona-se a postura intencional. A busca de uma explicação para o comportamento, por outro lado, pode continuar se reduzirmos o nível de análise para as posturas de projeto ou física.

In extreme cases personalities may prove to be so unpredictable from the intentional stance that we abandon it, and if we have accumulated a lot of evidence in the meanwhile about the nature of response patterns in the individual, we may find that a species of design stance can be effectively adopted. This is the fundamentally different attitude we occasionally adopt toward the insane. (Dennett, 1978a, p. 10)

Assim, finalizam-se os passos do método para a aplicação da postura intencional da seguinte maneira:

- 5) Observação do comportamento e avaliação da compatibilidade com os comportamentos pressupostos: confirmação ou segue-se para o passo 7.
- 6) Reformulação da previsão – voltar ao passo 4.
- 7) Reformulação da atribuição de crenças e desejos – voltar ao passo 2.
- 8) Abandono ou redução do nível de explicação.

Juntando todos os passos de um modo simplificado, temos o método a ser aplicado na próxima seção.

- 1) Especificação do contexto
 - a. História e seu lugar no mundo
- 2) Nível de atribuição
 - a. Desejos:
 1. Os meios mais praticáveis de satisfazer as necessidades biológicas básicas
 2. Atribuição de desejos específicos
 - b. Crenças:
 1. Atribuição de verdades relevantes disponibilizadas pela percepção do sistema
 2. Atribuição de crenças específicas
- 3) Especificação do comportamento
- 4) Nível de previsão
 - a. Explicação.
 1. Comportamento 1:
 - Crença inicial 1:
 - Crença inicial 2:
 - ...
 - Crença racional 1:
 - Crença racional 2:
 - ...

- Crença intencional:
 - Desejo intencional:
2. Comportamento 2:
 - ...
- b. Previsão:**
1. Hipótese 1:
 - Crença inicial 1:
 - Crença inicial 2:
 - ...
 - Crença racional 1:
 - Crença racional 2:
 - ...
 - Crença intencional
 - Desejo intencional:
 2. Hipótese 2:
 - ...
- 5) Verificação: confirmação ou refutação das hipóteses
 - 6) Reformulação da previsão
 - 7) Reformulação da atribuição
 - 8) Abandono e redução do nível de explicação

4.4 Aplicação para casos particulares

Nesta seção, utilizo o método concebido para a aplicação da postura intencional a casos específicos, com o objetivo de esclarecer as capacidades e limitações da teoria para lidar com casos de irracionalidade. Emprego, primeiramente, os casos paradigmáticos de atribuição de crença utilizados no início do capítulo para elucidar o funcionamento básico do método formulado. Em seguida aplico o método para um caso mais simples e para um caso mais complexo de irracionalidade, situações de falhas de raciocínio e delírio. Para a situação de falha de racionalidade utilizo o exemplo de Tversky e Kahneman (1983) usado no início do capítulo

e para o caso de delírio utilizo um diagnóstico clássico na literatura filosófica sobre delírios, o delírio de Capgras.

O que busco nestes exemplos é investigar se a TSI é capaz de lidar adequadamente com os casos de irracionalidade. O que seria lidar adequadamente? A prova para saber se a TSI consegue lidar adequadamente com os casos propostos seria se ela consegue dar conta de explicar tudo o que a PSC explica, ou seja, que tivesse, no mínimo, o mesmo poder explicativo da PSC. Se a TSI conseguisse mostrar isso, ela seria capaz de fornecer uma redução do mental a um domínio com entidades claramente definidas, cujos princípios de organização são familiares, relativamente formais e sistemáticos (Dennett, 1978, p. xviii).

4.4.1 Caso paradigmático

Empregarei o método formulado nesta seção a casos paradigmáticos de atribuição de crenças e desejos. Estes são casos em que há um consenso a respeito da atribuição de crenças e desejos, isto é, casos em que crenças e desejos determinados são atribuídos a uma pessoa, ou a um sistema, inequivocamente. Exemplos utilizados neste trabalho foram a mulher que usa o seu guarda-chuva para não se molhar e o caso de Stewart, que veste o seu chapéu antes de sair de casa para se proteger do frio. Estes casos são notáveis pela sua simplicidade e pela pequena quantidade de informação que é necessária para a sua interpretação. Utilizarei cada um desses casos para ilustrar as duas possibilidades de uso da TSI: a previsão e a explicação do comportamento.

O primeiro caso é o da mulher que utiliza seu guarda-chuva para não se molhar. Como utilizarei o caso para a previsão de comportamento somente, não passarei pelos passos 3 e 4a, pois estes são utilizados para a explicação do comportamento.

- 1) Especificação do contexto
 - a. Sua história e seu lugar no mundo
 - *Mulher adulta. Mora em uma cidade no mundo moderno. Se encontra em um dia chuvoso.*
- 2) Nível de atribuição
 - a. Desejos
 - Os meios mais praticáveis de satisfazer as suas necessidades básicas

- *Ela deseja segurança, saúde, ausência de dor ou sofrimento, sexo, alimentação, entretenimento, status, etc.*
- *Ela deseja não passar frio ou calor*
- *Ela deseja não se molhar para não passar frio*
- Atribuição de desejos específicos:
 - *Ela deseja andar pela rua*
- b. Crenças**
 - Atribuição de verdades relevantes disponibilizadas pela percepção do sistema
 - *Ela acredita que está chovendo e tem um guarda-chuva em suas mãos*
 - Atribuição de crenças específicas
- 4) Nível de previsão**
 - b. Previsão**
 - 1. Hipótese 1: *Abrir o guarda-chuva***
 - Crença intencional: *Ela acredita que se ela utilizar este guarda-chuva, não se molhará*
 - Desejo intencional: *Ela deseja não se molhar*
- 5) Verificação: confirmação ou refutação das hipóteses**
 - a. *Observa-se que a mulher abre o seu guarda-chuva e o utiliza para se proteger da chuva, confirmando a hipótese 1***

Como o que foi observado do seu comportamento é compatível com a previsão realizada, não há necessidade de reformulação da hipótese e a interpretação é considerada bem-sucedida.

Utilizaremos, agora, o caso de Stewart para a aplicação do método para a explicação de seu comportamento. Como não utilizaremos para a previsão de nenhum comportamento, então passarei ao largo os passos 5 e 6.

- 1) Especificação do contexto**
 - a. Sua história e do seu lugar no mundo**
 - 1. *Homem adulto. Está em um prédio do mundo moderno em um dia ventoso de inverno***
- 2) Nível de atribuição**
 - a. Desejos**
 - 1. Os meios mais praticáveis de satisfazer as suas necessidades básicas**

- *Ele deseja segurança, saúde, ausência de dor ou sofrimento, sexo, alimentação, entretenimento, status, etc.*
 - *Ele deseja ficar em casa para se proteger do frio.*
2. Atribuição de desejos específicos
 - *Ele deseja de sair de casa.*
- b. Crenças
1. Atribuição de verdades relevantes disponibilizadas pela percepção do sistema
 - *Ele acredita que está em casa em um dia ventoso de inverno.*
 - *Ele acredita que há um chapéu em casa.*
 2. Atribuição de crenças específicas:
 - *Ele acredita que este chapéu lhe protege contra o vento.*
- 3) Especificação do comportamento a ser analisado
 - a. *Pegou o seu chapéu e o vestiu antes de sair de casa.*
 - 4) Nível de previsão
 - a. Explicação
 1. Comportamento 1: *vestir o chapéu e sair de casa.*
 - Crença inicial 1: *Ele acredita que está ventando fora de casa.*
 - Crença inicial 2: *Ele acredita que o chapéu lhe protegerá do frio.*
 - Crença intencional: *Acredita que o chapéu lhe protegerá do frio se vesti-lo antes de sair de casa.*
 - Desejo intencional: *Se proteger do frio e de sair de casa.*
 - 7) Confirmação ou reformulação da atribuição de crenças e desejos – voltar ao passo 2
 - 8) Confirmação da hipótese ou abandono e redução do nível de explicação

Pode-se perceber facilmente que com os dois casos não há maiores dificuldades de interpretação pela postura intencional. Como não há problemas na explicação ou previsão do comportamento, a interpretação cessa até que novas motivações surjam para uma nova interpretação.

4.4.2 Falha de raciocínio

Nesta seção, aplicarei o método formulado para uma situação de falha de raciocínio comum. O exemplo utilizado vem do texto de Bortolotti ‘Intentionality without rationality’ (2005). Emprego este exemplo porque Bortolotti o utiliza como um modelo que exemplifica situações

de irracionalidade cotidiana e ordinária, e também porque ele aborda a mesma situação de irracionalidade analisada no início do capítulo.

O exemplo é o seguinte: Júlia, uma estudante de psicologia, se voluntaria para participar de um experimento sobre raciocínio. Este experimento é o mesmo que o aplicado por Tversky e Kahneman (1983) utilizado na introdução do presente capítulo. Digamos que Júlia faz parte dos 85% de pessoas que afirmaram que Linda, a pessoa descrita na pesquisa, tem uma menor probabilidade de ser uma caixa de banco do que ser, ambos, caixa de banco e ativa no movimento feminista. Ela comete, assim, a falácia da conjunção, a falha de raciocínio básica que as pessoas fazem quando tomam uma conjunção de dois eventos como mais provável do que algum dos eventos, isoladamente (Bortolotti, 2005, p. 388). Aplicando-se os dois primeiros passos para este exemplo temos o seguinte:

1) Especificação do contexto

a. Sua história e do seu lugar no mundo

- *Estudante de faculdade de uma cidade nos Estados Unidos da América no ano de 2014 responde a uma pesquisa dentro do campus a respeito de um problema de probabilidade.*

2) Nível de atribuição

a. Desejos

- Os meios mais praticáveis de satisfazer as suas necessidades básicas
 - *Desejo de segurança, saúde, sexo, alimentação, entretenimento, status, etc.*
- Atribuição de desejos específicos
 - *O desejo de escolher a alternativa correta em uma pesquisa é o esperado de estudantes.*

b. Crenças

- Atribuição de verdades relevantes disponibilizadas pela percepção do sistema
 - *Júlia acredita que está fazendo uma pesquisa sobre falhas de raciocínio.*
- Atribuição de crenças específicas
 - Sem crenças específicas disponibilizadas pelo exemplo.

O interessante desse exemplo é que não importaria se soubéssemos qual é o posicionamento de Júlia com relação à probabilidade de Linda ser uma caixa de banco, uma pessoa ativa no movimento feminista ou ambas as coisas: nenhuma aplicação da postura intencional lhe atribuiria a crença racional (ou seja, a crença derivada) de que ela acredita que a chance de

Linda ser C e F é maior do que a chance de ela ser só C. Assim, não importaria qual fosse a sua crença inicial, pois a sua ‘crença racional’ seria sempre a mesma, fazendo com que a previsão de seu comportamento também fosse a mesma. Como não é este comportamento que observamos, há a necessidade de reformular a atribuição de crenças e desejos para que a interpretação pela postura intencional funcione. Analisemos algumas hipóteses alternativas.

3) Especificação do comportamento

- a. Júlia responde à pesquisa de maneira a avaliar a probabilidade de Linda ser caixa de banco e ativa no movimento feminista como maior do que de Linda ser caixa de banco.

4) Nível de previsão

a. Explicação

- Hipótese 1:
 - Crença racional: *Júlia acredita que a chance de Linda ser C é maior ou igual à chance de Linda ser C e F juntas.*
 - Crença intencional: *Júlia acredita que, se ela responder que a chance de Linda ser C e F é maior que a chance de ser C, responderá incorretamente.*
 - Desejo intencional: *Ela deseja responder incorretamente à pergunta na pesquisa.*

Esta alternativa explicaria adequadamente o comportamento de Júlia. Pode-se verificar, contudo, que ela não é adequada, pois é incompatível com o desejo atribuído inicialmente de, na verdade, querer responder corretamente à pergunta da pesquisa. Noto que se a história de Júlia fosse diferente seria perfeitamente possível que essa hipótese explicasse o seu comportamento. Por exemplo, se ela soubesse que era o professor que a reprovou em uma matéria que estivesse fazendo a pesquisa, talvez ela desejasse retaliação e acreditasse que, se ela respondesse de modo que incorreto, enviesaria a pesquisa desse professor e dificultaria o seu sucesso profissional. Presumo, no entanto, que esta história, apesar de possível, dificilmente aconteceria com 85% dos participantes. Há a necessidade, portanto, de buscar por hipóteses alternativas de explicação do comportamento.

Outra hipótese viável é que Júlia tivesse uma crença de que a maneira adequada de calcular probabilidades seria a partir da similaridade dos perfis analisados. Então ela acreditaria em algo como: “É mais provável que Linda seja uma caixa bancária feminista do que uma caixa de banco, porque se parece mais com uma feminista ativa do que com uma caixa de banco”.

a. Hipótese 2:

- Crença inicial 1: *Júlia acredita que o mais importante para calcular a probabilidade de certas características da vida de uma pessoa é o que mais se assemelha ao perfil desta pessoa*
- Crença inicial 2: *Júlia acredita que Linda se parece mais com C e F do que com C*
- Crença intencional: *Júlia acredita que a resposta correta para a pergunta é que a probabilidade de Linda ser C e F é maior do que Linda ser C*
- Desejo intencional: *Júlia tem o desejo de responder corretamente à pesquisa*

Esta hipótese também explica adequadamente o comportamento de Júlia, só que sem violar a consistência de nenhuma das outras crenças ou desejos conferidos anteriormente. Se soubéssemos mais do histórico de Júlia, sabendo se ela é ingênua com relação às práticas formais básicas de lógica, como uma estudante nos primeiros anos de faculdade, esta hipótese seria mais provável. No entanto, se ela fosse uma estudante de doutorado em áreas voltadas para conhecimento estatístico com matérias avançadas em probabilidade, estatística e teoria da decisão, teríamos mais resistência em atribuir a ela esta crença e sustentar esta hipótese.

Aproveito esta hipótese para fazer uma digressão a respeito de um aparente problema com esta alternativa. Presumindo que Júlia tem uma opinião sobre a probabilidade de Linda ser C ou de ser F, a conclusão, aplicando-se uma racionalidade ideal, seria que ela sempre acreditaria que a chance de Linda ser C será maior ou igual que a chance de Linda ser C e F. Portanto, ao atribuímos a ela essa crença inicial 1 (*que o mais importante para calcular a probabilidade de certas características da vida de uma pessoa é o que mais se assemelha ao perfil desta pessoa*) e a subsequente crença racional (*que a resposta correta para a pergunta é que a probabilidade de Linda ser C e F é maior do que Linda ser C*), não estaríamos assumindo uma brecha de irracionalidade no nível de previsão por serem crenças incompatíveis?

A chave para entender este problema pela TSI está em diferenciar a crença em alguma coisa da decisão sobre a verdade de alguma sentença sobre esta coisa. Dennett (1987) denominaria este último de *opinião*, um ato de assentimento sobre a verdade de alguma sentença. Assim, ele diferencia marcadamente crenças de opiniões definindo opiniões como atitudes, comportamentos. Para ele, o ato de se apostar na verdade de alguma sentença não é um caminho exclusivo para uma crença com este conteúdo. Afinal, não é por consentirmos com

a conclusão de algum argumento, porque não conseguimos ver falhas em seus passos que, efetivamente, acreditamos naquilo.

Desse modo, a TSI diferencia a crença de alguém a respeito da melhor maneira de se raciocinar e o próprio modo de raciocinar da pessoa. Note que a crença que levou Júlia a ter o seu comportamento de responder à pesquisa não foi que ela acredita que “a probabilidade de Linda ser C e F é maior do que Linda ser C”, mas de que ela acredita que “*a resposta correta para a pergunta seria de que a probabilidade de Linda ser C e F é maior do que Linda ser C*”. Um exemplo que seria mais facilmente encontrado na mesma situação seria um fumante responder a respeito das chances de desenvolver algum tipo de câncer por fumar seria alta, mas continuar fumando, de modo que se ele *realmente* acreditasse nesta estimativa, quer dizer, se esta estimativa realmente motivasse suas disposições comportamentais, ele não fumaria (Dennett, 1978a, p. 308).

Uma questão que pode ter permanecido é a seguinte. Se, em alguma circunstância, as crenças de Júlia a respeito das probabilidades de Linda ser C ou F fossem cruciais para a tomada de alguma decisão (não só para uma resposta verbal para algum questionamento), não haveria a possibilidade de que Júlia, efetivamente, acreditasse ao mesmo tempo nas crenças contraditórias que a fizeram cometer a falácia da conjunção? Uma maneira de responder a esta questão é dizer que a falha em possuir a crença ‘a chance de Linda ser C não é menor do que a chance de Linda ser C e F’ como consequência lógica, assim como qualquer outra consequência lógica das crenças de um sistema, colocaria em dúvida as próprias crenças atribuídas, especialmente quando entram em contradição com outras crenças da rede. Isto resultaria na necessidade da reformulação da atribuição ou na necessidade de alguma história que justificasse o comportamento observado. É salutar para a validade da teoria de Dennett que é justamente esta atitude encontrada na formulação das hipóteses para estes comportamentos aparentemente irracionais no artigo de Tversky e Kahneman.

Outra hipótese que tem o potencial de explicar o comportamento de Júlia adequadamente é que ela, ao ler a afirmação a ser avaliada de quais seriam as chances de Linda ser uma caixa de banco, teria entendido a afirmação de uma maneira exclusiva, como se estivesse afirmando que Linda *só* seria uma caixa de banco e *nada mais*, ou seja, que seria uma caixa de banco e *não* uma pessoa ativa no movimento feminista. A questão, então, seria a respeito do que é mais provável caso Linda fosse C, se ela seria F ou não.

b. Hipótese 3:

- Crença inicial 1: *Júlia acredita que a pergunta a ser respondida é a respeito da probabilidade de Linda ser C e não ser F*
- Crença inicial 2: *Júlia acredita que, Linda sendo C, é mais provável que ela seja F do que não F*
- Crença racional: *Júlia acredita que a probabilidade de Linda ser C e F é maior do que Linda ser C e não ser F*
- Desejo: *Júlia deseja responder corretamente à pesquisa*
- Comportamento: *Júlia responde que a chance de Linda ser C e F é maior do que a chance de ser C*

Agora temos algumas alternativas para que possamos entender o comportamento de Júlia e os resultados do estudo de Tversky e Kahneman pela TSI. Ao mudarmos as crenças e desejos atribuídos inicialmente ao sistema, mesmo preservando a sua racionalidade, pudemos encontrar algumas alternativas para a explicação do comportamento “irracional” de Júlia.

Avaliando os resultados encontrados

Para avaliar os resultados encontrados, retomo a discussão a respeito de seus objetivos. Como foi visto, uma maneira crucial de avaliar a validade da TSI é comparando-se o seu poder explicativo e preditivo ao poder explicativa e preditivo da PSC (Dennett, 1987, p. 47). Como a TSI se propõe a ser o que tudo o que a PSC é em sua melhor forma, tudo o que a PSC for capaz de explicar, a TSI também deveria dar conta. Infelizmente não há um acesso fácil ao que é precisamente realizado dentro do uso comum da PSC e nem ao que exatamente evidenciaria o seu sucesso explicativo ou preditivo. Podemos comparar, no entanto, com outras formas de explicação do fenômeno analisado, que poderiam estar se utilizando de alguma forma da PSC para ajudar em suas formulações. Comparo, portanto, com as respostas dadas pelos próprios autores do artigo. Se a TSI puder dar conta de acomodar o que há de útil no modelo explicativo presente no artigo, seria um bom indicativo de que a TSI possui a validade teórica necessária e a viabilidade para ser incorporada pela ciência. Para este fim, uma comparação interessante é que as três hipóteses propostas neste exercício de aplicação da postura intencional também foram propostas pelos autores da pesquisa a respeito de Linda como possíveis explicações para o comportamento das pessoas pesquisadas.

Retomo a minha primeira hipótese. Ela sugere que haveria uma mudança substancial no nosso entendimento a respeito do desejo de Júlia em responder à pesquisa e que, por isso, ela

teria respondido de maneira incorreta. O que, no entanto, foi descartado devido a incompatibilidade com os desejos anteriormente atribuídos à Júlia. É interessante notar que Tversky e Kahneman também utilizam esta hipótese de que fatores motivacionais para acertar corretamente ao questionário pudessem estar envolvidos nos resultados encontrados. No entanto, como imaginado para Júlia, nenhuma mudança significativa nos resultados obtidos foi encontrada ao se adicionarem recompensas ou punições pelo acerto ou erro das questões, respectivamente (Tversky e Kahneman, 1973).

A última hipótese aventada acima diz respeito a uma dificuldade ou um erro de interpretação possivelmente realizados por Júlia. O equívoco seria em entender a afirmação de que ‘Linda é C’ como ‘Linda sendo exclusivamente C’, ou, no caso, C e não F. Por conta disso, ela teria crenças iniciais diferentes sobre o problema, que a levaria a respondê-lo de maneira equivocada. Uma hipótese equivalente é levantada no artigo de Tversky e Kahneman. Para testá-la, os autores realizaram uma pesquisa similar à primeira, mas que ressaltasse o sentido desta afirmação do seguinte modo: em vez de escreverem “Linda é uma caixa de banco”, eles escreveram, “Linda é uma caixa de banco, quer ela seja ou não ativa no movimento feminista” (Tversky e Kahneman, 1983, p. 17, tradução minha). A mudança nos resultados foi substancial o suficiente para concluir que este era um fator relevante dentre os que afetaram o resultado anteriormente encontrado, mas que a mudança de 85% para 57% não foi o suficiente para inverter a proporção de pessoas que aparentemente cometem a falácia da conjunção em suas respostas.

Por último, similarmente à segunda hipótese formulada no exemplo acima, considerou-se no artigo uma explicação que foca no conhecimento de Júlia sobre o problema em questão. Essa hipótese foca nas crenças de Júlia a respeito da melhor maneira de se resolver o problema em questão. Se Júlia acredita que a melhor maneira de descobrir a probabilidade de um evento é pensando em sua representatividade, então isso afetará as suas crenças sobre qual é a resposta adequada para o problema. Tversky e Kahneman também apostam nesta hipótese como um dos possíveis fatores envolvidos nos resultados encontrados. Para testá-la, eles aplicaram a pesquisa para uma amostra de estudantes de ciências sociais escolhidos por terem cursado diversas matérias sobre estatística. Com esta nova pesquisa, eles conseguiram uma margem ainda maior de pessoas que se conformaram à lei da conjunção, sendo que somente 36% dos participantes cometeram o erro.

A conclusão de que crenças sobre o modo de resolver adequadamente este problema alterariam a resposta para ele é corroborada também por outro estudo dos autores. Eles fizeram a mesma pergunta padrão a respeito de Linda para estudantes do primeiro ano em conjunto com

um segundo questionamento, em que se pedia para decidir qual de dois argumentos era o mais convincente:

Argument 1. Linda is more likely to be a bank teller than she is to be a feminist bank teller, because every feminist bank teller is a bank teller, but some women bank tellers are not feminists, and Linda could be one of them.

Argument 2. Linda is more likely to be a feminist bank teller than she is likely to be a bank teller, because she resembles an active feminist more than she resembles a bank teller. (Tversky e Kahneman, 1983, p. 16)

Uma parcela de 65% dos estudantes de graduação que responderam à pesquisa optaram pela alternativa incorreta, o argumento 2, como a estratégia que responderia de maneira adequada a questão da probabilidade. Estas duas pesquisas juntas mostram que o aprendizado cumpre um papel importante na correta avaliação de probabilidade das pessoas. Esta conclusão se encaixa adequadamente com a teoria de Dennett justamente por ela considerar que chegar a um consentimento, ou em uma aposta para a verdade de alguma sentença, como a respeito de Linda, é um comportamento intencional e que, portanto, pode ser modificado pela alteração de desejos e crenças a respeito do melhor jeito para se chegar a estes desejos.

A TSI, portanto, dá conta de formular adequadamente as principais hipóteses alternativas elaboradas pelos autores do artigo que justificariam o comportamento aparentemente irracional dos sujeitos de pesquisa. Este fato é potencialmente relevante pois, já que estas hipóteses não fazem parte da teoria psicológica particular dos autores ou de outra escola psicológica, presumo que deve haver alguma correlação da formulação envolvida nestas hipóteses com a formulação que a PSC produziria.

Um outro desafio para a TSI seria dar conta de acomodar em seu modelo teórico a própria teoria formulada por Tversky e Kahneman. Essa teoria diz respeito à utilização de processos heurísticos de avaliações naturais automáticas considerada, pelos autores, como estrutura teórica mais adequada para fundamentar as explicações dos resultados encontrados em suas pesquisas. Segundo Dennett, a proposta dos autores se encaixaria no que ele chama de psicologia sub-pessoal, que é a ciência sobre os reais funcionamentos dos sistemas intencionais. A teoria da psicologia sub-pessoal poderia avançar no seu conhecimento sobre o funcionamento do cérebro humano, obscurecendo as fronteiras com a TSI, assim como hoje é obscurecida a fronteira entre a física e a química. Todavia, o objetivo final da psicologia sub-pessoal deve ser demonstrar como um sistema descrito em termos psicológicos pode merecer uma interpretação em termos intencionais, pois qualquer interpretação deve, inevitavelmente, depender da integralidade das crenças e dos desejos do sistema.

Psychology can, and should, work toward an account of the physiological foundations of psychological processes, not by eliminating psychological or intentional characterizations of those processes, but by exhibiting how the brain implements the intentionally characterized performance specifications of sub-personal theories. (Dennett, 1987, p. 64)

A proposta de Dennett é, portanto, que os avanços teóricos da ciência psicológica como proposto no artigo de Tversky e Kahneman podem ajudar na interpretação intencional de seres humano só na medida em que validam teoricamente a aplicação da TSI.

There is also no doubt a design stance explanation at a somewhat higher level of abstraction of the chess player who is suckered by knight forks – there is a sub-optimal bias in his heuristic search-tree – pruning dispositions, for instance. I never deny that we can take advantage of design stance discoveries; I just insist that they are interpreted as design stance implementations of intentional stance competences, and those competences are still describable only in rationality-presupposing language. (Dennett, 1993a)

A minha proposta aqui é que a forma que a TSI acomoda esta legitimação, incremento e correção do seu modo de funcionar é pela atribuição diferenciada de crenças no nível de atribuição. Ilustro aqui, *grosso modo*, como ficaria esta interação da psicologia sub-pessoal e da TSI no caso da heurística da representatividade.

Na heurística da representatividade, a probabilidade de que uma alguma coisa pertença a um grupo é medida pela similaridade das características mais marcantes de membros desse grupo com as características desta coisa (Tversky e Kahneman, 1983). A formulação deste processo heurístico está inserida na estrutura teórica da psicologia cognitiva, que lida com os fenômenos mentais em termos de processamento de informação e funções cognitivas, como memória, atenção, percepção e, no caso, o raciocínio (Sternberg & Sternberg, 2011). Em um nível de projeto, pode-se entender o funcionamento das pessoas, mas para utilizar isso na prática para prever o seu comportamento em nível de intencionalidade, é preciso codificar o conhecimento adquirido em termos de crenças e desejos particulares para o sistema.

Portanto, uma regra geral para a atribuição de crenças a pessoas, incluindo Júlia, pode ser estabelecida no nível de atribuição: as pessoas tendem a acreditar que a probabilidade de um evento é similar à probabilidade de um outro evento do qual este primeiro é representativo. Então, o que fazemos é atribuir a Júlia uma crença particular ao pensar sobre o exemplo.

- 1) Especificação do contexto.
 - a. História e seu lugar no mundo.

1. *Estudante de faculdade de uma cidade nos Estados Unidos da América no ano de 2014 responde a uma pesquisa dentro do campus a respeito de um problema de probabilidade.*
 2. *Júlia reflete a respeito das características mais prováveis sobre Linda*
- 2) Nível de Atribuição:
- a. Desejos:
 1. Os meios mais praticáveis de satisfazer as necessidades biológicas básicas.
 - *Desejo de segurança, saúde, sexo, alimentação, entretenimento, status, etc.*
 2. Atribuição de desejos específicos.
 - *O desejo de escolher a alternativa correta em uma pesquisa é o esperado de estudantes.*
 - b. Crenças:
 1. Atribuição de verdades relevantes disponibilizadas pela percepção do sistema.
 - *Júlia acredita que está fazendo uma pesquisa sobre falhas de raciocínio.*
 - *Júlia acredita que o perfil de Linda é representativo de uma pessoa que seja F do que de uma pessoa que seja C.*
 2. Atribuição de crenças específicas.
 - *Pela regra empírica aprendida: as pessoas acreditam que a probabilidade de um evento é similar à probabilidade de um outro evento do qual este primeiro é representativo. Júlia acredita é mais provável que Linda seja F e C do que C.*

A partir desta atribuição de crenças que podemos entender e prever as suas respostas para a pesquisa intencionalmente.

- 3) Especificação do comportamento.
 - a. Júlia responde à pesquisa de maneira a avaliar a probabilidade de Linda ser caixa de banco e ativa no movimento feminista como maior do que de Linda ser caixa de banco.
- 4) Nível de previsão.
 - a. Explicação.

1. Comportamento 1: avaliar que a probabilidade de Linda ser C e F é maior do que Linda ser C
 - Crença inicial e intencional: *Júlia acredita que a chance de Linda ser C e F é maior do que a chance de Linda ser C.*
 - Desejo inicial e intencional: *Ela deseja responder corretamente à pergunta na pesquisa.*

4.4.3 Delírio de Capgras

Nesta seção, realizarei a aplicação do método formulado e uma posterior análise dos resultados para um caso de uma pessoa com uma condição psicótica denominada delírio de Capgras. O exemplo utilizado vem do artigo de David Rose, Wesley Buckwalter e John Turri (2014), que utiliza um exemplo de uma pessoa com este delírio para pesquisar qual seria a visão da PSC de casos de delírio referente aos seus estados mentais, isto é, se estes estados seriam caracterizados pela PSC como crenças (ou estados imaginativos, por exemplo). Primeiramente, faço uma introdução ao delírio de Capgras com uma breve descrição de suas características principais, para obter a compreensão apropriada para a discussão.

Segundo Bortolotti (2005), pessoas sujeitas à síndrome de Capgras alegam que seus cônjuges, ou um de seus parentes próximos, foram substituídos por impostores. Segundo uma teoria reputada (Ellis *et al.*, 1997), esse delírio seria causado por um dano no módulo de reconhecimento facial em que a pessoa vê o cônjuge, mas não o reconhece. Além de uma experiência anormal, um viés de raciocínio também é necessário para a aceitação e manutenção do delírio (Davies *et al.* 2001). Em alguns casos, os pacientes de Capgras agem com base em suas crenças delirantes, mostrando um comportamento hostil ou agressivo em relação aos supostos impostores. Quanto mais complexas e articuladas as explicações dadas pelos sujeitos, mais elaborado é o seu estado delirante, onde a elaboração é medida em termos do número de conexões que o estado delirante tem com outras crenças possuídas pelo sujeito. Alternativamente, os sujeitos podem continuar com suas vidas substancialmente inalteradas e mostrar muito pouca preocupação sobre sua situação incomum e sobre o suposto desaparecimento de seus cônjuges ou parentes próximos. Nestes últimos casos, o delírio é circunscrito, isto é, o estado delirante não necessariamente interage com as outras crenças, emoções e desejos do paciente e não necessariamente leva o sujeito a agir de acordo com ele, provavelmente como um efeito de alguma forma de compartimentação. Uma característica geral dos delírios elaborados e circunscritos parece ser a perseverança com que os sujeitos

mantêm o delírio, mesmo frente a análise do conteúdo implausível de seus relatos e dos esforços de outras pessoas para dissuadi-los. Como afirma Bortolotti, a resistência à mudança e a compartimentação afetam a capacidade do sujeito de avaliar e rejeitar seu estado delirante, indica que há um elemento de irracionalidade no compromisso do sujeito com o conteúdo do seu delírio (2005, p. 191-192)

Utilizo o caso proposto para pesquisa de Rose *et al.* (2014). A descrição inicial do caso se desenvolve da seguinte maneira:

1) Especificação do contexto e do comportamento do sistema

a. Constatação da sua história e do seu lugar no mundo

- Don e Katherine estavam casados há dez anos. Como a maioria dos casais, eles passaram muito tempo juntos e fizeram muitas coisas juntos. Eles regularmente compartilhavam as refeições, iam ao cinema, moravam na mesma casa e dormiam na mesma cama. Então, um dia, quando Don estava dirigindo-se para a loja, um carro furou o sinal vermelho, batendo na porta do motorista do seu carro. Don sofreu uma lesão cerebral traumática como resultado do horrível acidente de trânsito. No ano seguinte de sua lesão, ele começou o que parecia ser uma recuperação notável e rápida. Don recuperou seus poderes de fala, inteligência e quase todas as suas habilidades cognitivas e sociais. (Rose *et al.*, 2014, p. 686)

Analisemos como uma pessoa comum, utilizando somente o que é relevante da PSC e seus conhecimentos gerais interpretaria um caso clássico na literatura sobre delírios como o delírio de Capgras. Com o conhecimento de uma pessoa comum e o modo de aplicação formulado da postura intencional, pressuponho as seguintes atribuições de crenças e desejos:

2) Nível de atribuição

a. Desejos

- Os meios mais praticáveis de satisfazer as suas necessidades básicas
 - *Don deseja sobreviver, a ausência de dor, segurança, companhia, saúde, sexo, alimentação, conforto, poder, influência, entretenimento.*
 - *Don deseja se recuperar o mais rápido possível para que deixe de sofrer com dor, não quer voltar ao hospital e deseja ficar no conforto de sua casa, etc.*

- Atribuição de desejos específicos
 - *Don deseja viver com sua esposa e ter a sua companhia.*
 - *Don deseja o bem estar de sua esposa*

b. Crenças

- Atribuição de verdades relevantes disponibilizadas pela percepção do sistema
 - *Don acredita que é casado com a sua mulher, sofreu um acidente do qual teve uma recuperação quase total, mora em sua casa, etc.*
 - *Don acredita que está em sua casa com a sua mulher, etc.*
 - *Ele acredita naquilo que professa, em circunstâncias em que não é coagido*
 - *Don acredita que, se a sua esposa desaparecesse, ele poderia procura-la ou chamar a polícia*

3) Especificação do comportamento a ser analisado

- Apesar de sua recuperação, há algo de muito estranho em seu comportamento. Don conta para seus amigos, familiares e médicos que a sua esposa teria sido substituída por uma impostora
- Ele não foge da sua suposta impostora; ele não chama a polícia; e ele não procura a sua querida esposa. Ele continua a compartilhar a sua vida com a suposta impostora, sem, mesmo, explosões violentas de comportamento

Para o próximo passo, entramos no nível de previsão ou explicação do comportamento. Quais seriam os possíveis pares de crenças e desejos que poderíamos atribuir a Don de modo a explicar seus comportamentos? Neste momento, se impõe uma marcada dificuldade para responder a essa questão, pois poucas crenças que explicariam o comportamento também seriam consistentes com as atribuições iniciais realizadas de que a) Don acredita naquilo que professa quando diz que ela é uma impostora, b) Don ama a sua esposa e se preocupa com ela e c) se preocupa com a sua segurança. Isto acontece porque, dentro das crenças atribuídas no passo 2, está a crença que Don acredita que pode fazer algo para buscar a segurança de sua esposa, seja procurando por ela com seus amigos ou ligando para a polícia.

Como vimos, aplicar a postura intencional é como montar um jogo de quebra-cabeça, em que as peças iniciais são montadas primeiro, as atribuições iniciais, e as peças que montam a imagem esquemática da explicação intencional são peças disponíveis para serem encaixadas. Achar uma explicação é encontrar peças que se encaixem com as todas peças iniciais, mas também com as peças que desvendam o comportamento. Se isto não é alcançado na

interpretação, algo deve ser renunciado. Ou reconhecemos, ao menos naquele momento, que a pessoa deixou de acreditar em algo que atribuímos a ela inicialmente, removendo a peça inicial que não se encaixa com o resto, ou abandonamos a imagem e a interpretação intencional. Desta maneira, segundo a TSI, o comportamento de Don de não procurar a sua esposa ou de não comunicar à polícia não é explicado enquanto pressupormos que ele deseja a sua companhia e segurança e que pode fazer algo para alcançar isto, seja ligando para a polícia ou através de outra atitude.

Um grande problema para a TSI ao lidar com delírios (mesmo delírios ditos *monotemáticos* como é o caso da síndrome de Capgras) é que as pessoas delirantes acabam verbalizando tantas crenças falsas que se torna laborioso, quando não impossível, montar uma rede de crenças coerente que devolva racionalidade para os seus comportamentos de forma que possam ser entendidos intencionalmente. E, quando se tenta devolver a racionalidade, perde-se boa parte da integridade das crenças verdadeiras que um sistema deve possuir no nível de atribuição de crenças. Consideremos um exemplo para esclarecer este ponto.

Uma hipótese para justificaria o comportamento de Don pela postura intencional poderia ser algo como o seguinte cenário. Don acredita que a sua esposa não está mais lá, foi abduzida e há uma impostora em seu lugar. Mas ele também acredita que sua esposa provavelmente está bem, em algum outro lugar, pois mandou para ele sinais criptografados por padrões de carros que passam em sua rua, comunicando que ela está bem e logo voltará. Ele também acredita que esta impostora não é só fisicamente parecida com a sua esposa, mas também é querida como ela e essa situação não é tão ruim assim.

5) Nível de previsão

a. Explicação

1. Comportamento 1: Contar a todos que sua esposa foi substituída por uma impostora

- Crença inicial 1: *Ele acredita que a sua esposa foi abduzida*
- Crença inicial 2: *Ele acredita que algumas pessoas são confiáveis que lhe ajudariam*
- Crença racional 1: *Se ele contar para estas pessoas que a sua esposa foi abduzida, elas poderiam ajudá-lo*
- Crença intencional: *Ele acredita que se ele contar para as pessoas, elas podem ajudá-lo a achar a sua esposa e lidar com impostora*

- Desejo intencional: *Ele tem o desejo de lidar com esta impostora e conseguir ajuda para achar sua esposa*
2. Comportamento 1: Não procurar a sua esposa ou ligar para a polícia
- Crença inicial 1: *Ele acredita que a sua esposa foi abduzida*
 - Crença inicial 2: *Ele acredita que sua esposa manda sinais periódicos de que está tudo bem e que logo voltará para casa*
 - Crença inicial 3: *A policial traz incômodo para a família das pessoas desaparecidas*
 - Crença racional 1: *Don acredita que, como ela já está vindo para casa, a polícia só traria incômodo*
 - Crença intencional: *Se ele procurar a polícia, só trará mais incômodo*
 - Desejo intencional: *Não se incomodar*
3. Comportamento 2: Viver pacificamente com a “impostora”
- Crença inicial 1: *Don acredita que esta pessoa igual à sua esposa é uma impostora*
 - Crença inicial 2: *Don acredita que, apesar de tudo, ela é uma pessoa agradável de se conviver*
 - Crença intencional: *Don acredita que, enquanto a sua esposa não volta, ele pode viver com esta pessoa para lhe fazer companhia*
 - Desejo intencional: *Don deseja companhia em sua vida*

Deixo claro o grande esforço que é encontrar crenças e desejos que justifiquem o comportamento de alguém que se afasta tanto do que é esperado ordinariamente. Mesmo nesta exótica atribuição de crenças, há uma contradição presente nas crenças invocadas para explicar os comportamentos. Como Don acredita que a sua esposa foi abduzida e precisa de ajuda, mas também acredita que ela está bem e por isso é melhor não chamar a polícia? Apresento este exemplo para esclarecer que esta explicação, apesar de justificar os comportamentos isoladamente, não se acomoda confortavelmente em nossa interpretação. Algo deve ser dito para resolver a contradição existente. Don não lembrava de certos fatos em alguns momentos mas se lembrava em outros? Ou, quem sabe, ele tem algum *outro* tipo de conjunto de crenças? Enquanto não houver respostas, a instabilidade e inadequação dominarão a interpretação.

A partir dessas atribuições de crenças iniciais podemos tentar entender intencionalmente os seus comportamentos de avisar que sua esposa teria sumido, não agir como se a sua esposa fosse ela mesma e também não procurar ajuda para a polícia. Muitas vezes, no entanto, nem há necessidade de o sujeito de interpretação engendrar hipóteses tão anômalas para devolver esta racionalidade procedural à pessoa interpretada, porque a própria pessoa delirante já faz isto para ele. Segundo Bortolotti (2005), quando pressionados em suas inconsistências, os pacientes com delírio de Capgras também podem formar outras crenças relacionadas ao conteúdo de seu estado delirante na tentativa de explicar como seu delírio se encaixa com outras coisas que eles conhecem. A autora relata que se pediu a um paciente que explicasse por que o “impostor” tinha o anel que ele deu à esposa e respondeu que o anel não era o mesmo, apenas um anel muito parecido. Igualmente, foi-lhe perguntado por que não havia relatado o desaparecimento de sua esposa à polícia e ele respondeu francamente que a polícia nunca teria acreditado nele. Apesar de muitas vezes serem duvidosas em sua veracidade, estas crenças viabilizam a interpretação pela postura intencional pois devolvem para ela a sua racionalidade para o nível de previsão.

Novamente, vemos que a coesão racional é essencial para a interpretação intencional no nível de previsão e explicação de comportamentos. Destaco uma passagem de Dennett que discorre a respeito disso com respeito a um caso de dissonância cognitiva.

But which of these is actually believed will be determined by seeing what he says and does, and then asking: what beliefs and desires would make those acts rational? And whatever delusion is embraced, it must be — and will be — carefully surrounded by plausible supporting material, generatable on the counterfactual assumption that the delusion is an entirely rationally held belief. Given what we already know about Jones, we might be able to predict which comforting delusion would be most attractive and efficient for him — that is, which would most easily cohere with the rest of the fabric of his beliefs. So even in a case of cognitive dissonance, where the beliefs we attribute are not optimal by anyone's lights, the test of rational coherence is the preponderant measure of our attributions. (Dennett, 1987, p. 100)

Todavia, como estas crenças delirantes apareceram em seus estados mentais é algo para a psicologia sub-pessoal (seja a psicologia cognitiva ou a neuropsicopatologia) responder. Quais os processos que estão por trás de crenças tão desconectadas com a realidade? As descobertas da psicologia sub-pessoal, no entanto, não poderão substituir a TSI se a intencionalidade pretende ser preservada. Elas apenas poderão legitimar a atribuição particulares de crenças, desejos ou outros estados mentais derivados. Há casos, no entanto, em que a irracionalidade é tamanha que, sem um entendimento mais aprofundado do funcionamento interno para auxiliar na atribuição de crenças, nenhuma crença pode ser atribuída à pessoa, e talvez algum tipo de explicação pela postura de projeto poderá ser mais adequada.

In extreme cases personalities may prove to be so unpredictable from the intentional stance that we abandon it, and if we have accumulated a lot of evidence in the meanwhile about the nature of response patterns in the individual, we may find that a species of design stance can be effectively adopted. This is the fundamentally different attitude we occasionally adopt toward the insane. To watch an asylum attendant manipulate an obsessively countersuggestive patient, for instance, is to watch something radically unlike normal interpersonal relations. (Dennett, 1978a, p. 10)

Retomo agora o artigo cujo exemplo utilizamos na presente análise. Nele, os autores (Rose *et al.*, 2014) buscaram encontrar qual seria a visão da PSC a respeito da atribuição de crenças a casos de delírio como o de Don. O artigo é de particular importância para o presente trabalho pois é um de dois artigos encontrados que buscam o entendimento da PSC a respeito de casos de delírios.⁴ Concordo com os autores em sua afirmação de que há uma carência sentida de dados como estes. “Para um debate amplamente centralizado na psicologia popular da crença, é surpreendente que não haja investigação empírica acerca da prática popular de atribuição de crenças em casos de delírio” (Rose *et al.*, 2014, p. 16).

Analiso, agora, a pertinência do artigo para a presente investigação. No artigo, os autores buscam testar a hipótese de que a PSC vê delírios como casos inequívocos de crenças estereotípicas. Para tanto, eles mostram algumas pesquisas que, de acordo os autores, apontam nitidamente para esta direção, chegando a três conclusões: 1) as pessoas prontamente classificam delírios como crenças; 2) as pessoas veem delírios como crenças pois a frequente asserção é um poderoso indicativo para a atribuição de crenças; e 3) pacientes delirantes são prontamente vistos como possuindo crenças contraditórias.

Estas parecem ser conclusões preocupantes para a TSI. Primeiramente, segundo o caso analisado de Don pela postura intencional, não seria tão rápida a atribuição de crenças, dado que esta análise apresenta acentuada complexidade, o que se contrapõe à primeira conclusão do artigo. Em segundo lugar, de acordo com a TSI, a frequente asserção pode ser um indicativo para a atribuição de crenças, mas está longe de ser o principal, que é a totalidade de seus comportamentos, o que diverge da segunda conclusão. Mas a mais condenatória das conclusões seria a terceira, pois a TSI jamais poderia prontamente admitir e atribuir crenças contraditórias a um sistema.

Poderíamos concluir que estas são evidências que demonstram uma falha na TSI, já que a PSC, por interpretar diferentemente estes casos, estaria claramente funcionando de um modo

⁴ Além do estudo mencionado, há um estudo subsequente (Rose *et al.*, 2017) a este que tem o propósito de ser uma reprodução da primeira pesquisa em escala mundial (i.e., transcultural). Todas as considerações feitas com relação ao primeiro artigo podem ser estendidas para o segundo.

distinto da dela? Este seria um caminho de investigação interessante, mas alguns passos ainda devem ser dados e algumas considerações realizadas antes de chegarmos a esta conclusão.

Primeiramente, não é o caso que a TSI se propõe a ser exatamente o que a PSC é, mas somente a respeito do que há de útil nela. Assim, para que a TSI possa ser invalidada em seu propósito, não só deve haver um distanciamento da TSI com relação à PSC, mas a PSC tem que ser capaz de explicar e prever eventos que não sejam explicados ou previstos pela TSI. Se os autores do artigo estão corretos em concluir que a PSC identifica delírios como casos típicos de crenças, uma forma de avaliar o poder explicativo e preditivo da PSC e posteriormente a viabilidade da TSI seria a partir dos seguintes passos:

- 1) Avaliar o quanto a PSC se distancia do uso da TSI para realizar estas previsões;
- 2) Se há um distanciamento, identificar qual seria a previsão pela TSI;
- 3) Aferir o quanto a caracterização de delírios como crenças típicas pode, de fato, explicar os comportamentos das pessoas delirantes;
- 4) Comparar a eficácia de cada previsão.

Possibilitar a realização do primeiro passo e do segundo passo é o propósito da construção do presente modelo de aplicação da TSI. Com o modelo de aplicação da TSI construído pode-se avaliar se uma interpretação está, ou não, de acordo com os seus princípios e funcionamentos adequados. O terceiro passo poderia ser realizado com uma análise da literatura empírica acerca dos delírios (o que, no entanto, está para além do escopo e da possibilidade de realização do presente trabalho).

Finalmente, ofereço um adendo sobre a conclusão de que o estudo de Rose *et al.* mostra que a PSC se afasta da TSI, a saber, que não é causa perdida que a TSI possa explicar os resultados obtidos na pesquisa. Dennett introduz o conceito de opinião para distinguir estados cognitivos que incorporam linguagem – apostas na verdade das sentenças – de estados cognitivos portadores de informação, como a crença de animais (Dennett, 2015, p. 208). Se a distinção entre crenças e opiniões for significativa, não seria inconcebível que as pessoas estejam utilizando um significado da palavra ‘crença’ de um modo similar ao conceito de opinião, especialmente pela falta de alternativas em nosso vocabulário para expressar tal conceito. Como uma das grandes diferenças de ‘opinião’ para ‘crença’ é a sua utilização exclusiva por sistemas possuidores de linguagem, uma questão que sondasse se a pessoa estaria

disposta a atribuir o mesmo estado que atribuiu à pessoa delirante a outros animais ajudaria a discernir o sentido utilizado.⁵

4.4 Conclusão

Para dar continuidade ao propósito de busca de uma resposta para a questão da viabilidade da TSI, o presente capítulo teve como objetivo a aplicação da TSI para casos de difícil atribuição de crenças por conta da irracionalidade apresentada. A fim de efetuar esta tarefa, realizei a construção de um método claro e rigoroso de aplicação da postura intencional para qualquer caso em que uma interpretação intencional se faça necessária. Dividi esta aplicação em dois níveis e delineei os seus funcionamentos dentro dela. Em seguida, expus algumas considerações importantes para o uso do método e o formalizei em uma sequência estruturada de passos para a sua utilização. Finalmente, apliquei o método para casos escolhidos de irracionalidade e analisei os resultados encontrados e as suas implicações para a validade da TSI.

Meus esforços iniciais desta empreitada foram no sentido de aplicar a TSI da maneira como foi formulada por Dennett. Encontrei, no entanto, uma dificuldade teórica, pois o método oferecido se mostrou rudimentar e insatisfatório para lidar com as complexidades encontradas na aplicação a casos com marcada irracionalidade. Assim, mesmo com os princípios norteadores do critério de racionalidade da TSI expostos, não houve uma clareza no modo que a TSI lida, na prática, com a irracionalidade.

Os dois níveis de aplicação então formulados foram concebidos com a intenção de superar esta dificuldade e de esclarecer a função que os princípios do critério de racionalidade têm dentro de sua aplicação. Enquanto que o primeiro nível, o nível de atribuição de crenças e desejos, tem a função de reduzir a indeterminação, o segundo nível, de previsão e explicação, tem o objetivo de sustentar a rede de crenças e realizar a ligação entre crenças, desejos e comportamentos.

⁵ Outra maneira de efetuar uma pesquisa pertinente para os propósitos de avaliar a viabilidade da TSI seria (1) identificar quais seriam os comportamentos previstos pela PSC para um caso de delírio; (2) identificar o modelo utilizado para se chegar a esta resposta; e (3) avaliar o êxito da sua previsão. Um exemplo da aplicação dessa proposta de pesquisa seria requerir que os participantes não meramente respondam a pergunta de o que explica o comportamento das pessoas delirantes (se crenças, a ausência delas, ou “pensamentos”, ou estados imaginativos), mas pedir para que prevejam qual seria o seu comportamento em determinada situação e então investigar quais teriam sido os motivos que as levaram a oferecer tais respostas. O levantamento das suas previsões e a visualização dos motivos relatados pelos sujeitos de pesquisa auxiliaria na averiguação de o que é compatível com os resultados de uma aplicação da postura intencional, obtida mediante o método esboçado. Uma comparação com os comportamentos tipicamente encontrados em casos de delírio poderia, então, oferecer uma pista a respeito da eficácia explicativa de cada uma de suas previsões.

Com relação ao nível de previsão e explicação de comportamentos, mesmo que tenha o mesmo objetivo da noção de racionalidade por inteligibilidade cunhada por Bortolotti, que seria a aplicação da razão prática visando uma interpretação intencional, vimos que a sua formulação desta noção não se identifica com todas as características que são basilares para aplicação da postura intencional. Três diferenças foram identificadas: a falta de um reconhecimento da necessidade da utilização de crenças implícitas; a ligação particular entre crenças, desejos e comportamentos delineada pela equação intencional; e um modelo particular de racionalidade que deve ser aplicado a qualquer interpretação intencional. Assim, concluí que é necessário, para a previsão ou explicação de comportamentos pela postura intencional, encontrar pares de crenças e desejos que satisfaçam a equação intencional. Esses pares de crenças e desejos devem ser encontrados em meio a uma rede holística de crenças, racionalmente estruturada e interligada.

Já o nível de atribuição tem a função particular de reduzir a indeterminação da interpretação. A TSI alcança esta redução por meio de uma pressuposição de que o sistema tenha, em sua maior parte, estratégias racionais e seja, em sua maior parte, projetado de maneira ótima. Vimos que Dennett defende a validade desta estratégia com o seu argumento evolutivo, que sustenta que estratégias racionais e crenças verdadeiras, para o que é mais importante para o organismo, serão encontrados em seres moldados pelo processo de evolução por seleção natural. O ponto chave que faz com que possamos reduzir a indeterminação da interpretação é a pressuposição de um projeto ótimo, o que faz com que possamos utilizar a TSI não só para organismos que evoluíram, mas também para artefatos criados por seres inteligentes. Todavia, dentro deste nível de aplicação, a pressuposição de um projeto ótimo não é o único método de descobertas de atribuição de crenças e desejos. Descobertas empíricas podem (e devem) auxiliar neste processo realizando os ajustes finos da atribuição, enquanto que a maior carga é levantada pela pressuposição de um projeto ótimo.

Assim, os dois níveis funcionam em conjunto para possibilitar a aplicação da postura intencional a qualquer sistema. Há, primeiramente, a aplicação do nível de atribuição, em que há atribuições iniciais de crenças e desejos. Em segundo lugar, emprega-se o segundo nível de aplicação, o nível de previsão, em que se explica o comportamento sob análise ou se prevê o comportamento futuro do sistema.

Ao avaliar a relação que a TSI permite entre crenças e desejos e emoções e sentimentos, cheguei a duas conclusões importantes. Primeiro, a busca de crenças e desejos adequados para a interpretação intencional pode ser auxiliada avaliando os efeitos que crenças e desejos podem ter no sistema, como emoções ou o enrubescer de faces, que não sejam unicamente os

comportamentos intencionais. Segundo, pode-se utilizar ambas as posturas de projeto e física para esta busca de atribuição inicial, seja através de emoções que gerem desejos particulares ou pela mera exposição a um evento que gere uma crença sobre ele.

Ao revisitar os argumentos levantados contra o critério de racionalidade para atribuição de crenças utilizado pela TSI, pôde-se esclarecer e reforçar a sua defesa diante de suas críticas. O argumento do duplo padrão de racionalidade pôde ser devidamente respondido ao se notar as utilizações distintas, mas complementares, da racionalidade para cada nível de aplicação. A defesa ao argumento da dissimilaridade com a PSC é corroborada esclarecendo-se como a TSI lida com situações de irracionalidade. Reanalizando o argumento do critério arbitrário da redução da indeterminação, conclui que há mérito ao afirmar que a pressuposição de racionalidade para a redução da indeterminação (i.e., de um projeto ótimo) não é estritamente necessária, mas o argumento se mostra indevido ao apontar a racionalidade como critério arbitrário, pois há justificativas contundentes para a utilização do critério como âncora para a atribuição de crenças e desejos. Já o argumento da definição arbitrária é respondido pela possibilidade de se produzir hipóteses independentes e testáveis, como evidenciam as seções finais deste trabalho.

A aplicação do método formulado para Júlia, o caso de falha de raciocínio dentro do experimento proposto por Kahneman e Tversky, possibilitou a observação de três resultados interessantes para os fins desta dissertação. Primeiro, todas as hipóteses alternativas sugeridas pelos autores do artigo puderam ser reproduzidas pela TSI como hipóteses, de modo que explicassem o comportamento da estudante apelando para as mesmas crenças e desejos. É particularmente relevante que a TSI tenha conseguido reproduzir estas explicações por dois motivos. Faz sentido esperar que parte das explicações que não têm suas origens em teorias psicológicas particulares tenham um grande impacto da PSC em suas formulações. O segundo motivo é que duas das três explicações analisadas foram corroboradas pelos resultados como explicações com certo poder explicativo. Assim, apesar de haver a necessidade de estudos com formulações mais claras a respeito do que seria precisamente explicações pela PSC, estes resultados não deixam de ser um indicativo de que a TSI é capaz de lidar com, ao menos, algumas explicações produzidas pela PSC com valor explicativo para casos de irracionalidade.

A segunda observação que foi possibilitada pela análise dos resultados é a constatação de que não foi necessário atribuir uma racionalidade abaixo do ideal no nível de previsão para chegar estes resultados, mas somente modificar as atribuições de crenças e desejos inicialmente atribuídos ao sistema, assim como conceitualizado na primeira parte do capítulo. E, terceiro, pode-se perceber de que modo a heurística de representatividade, que é uma formulação

particular dos autores do artigo sobre o funcionamento do aparato cognitivo humano, pôde ser adaptada para operar pela TSI, possibilitando uma interpretação intencional. Vimos que o que possibilita esta interação de explicações da postura de projeto e da postura intencional é a formulação de regras particulares de atribuição de crenças e desejos para circunstâncias específicas nas quais o sistema se encontra.

Já a aplicação para o caso de Don, o paciente diagnosticado com delírio de Capgras, não foi tão direta. Vimos que a dificuldade para a interpretação deste caso ocorreu pela dificuldade em construir uma rede de crenças que, ao mesmo tempo, explicasse o seu comportamento e que fosse íntegra e coerente. Vimos que crenças incomuns podem ser utilizadas para explicar certos comportamentos de pacientes delirantes, mas que elas geralmente não são facilmente acomodadas dentro da rede de crenças do indivíduo. Quando há a possibilidade de construir uma rede coerente de crenças e desejos que racionalmente expliquem o seu comportamento, pode-se atribuir as crenças sem maiores ressalvas. Mas, quando isto não acontece, uma instabilidade se apresenta e gera a exigência de uma explicação. Desse modo, apesar da postura intencional não produzir explicações em todas as situações para casos de delírio, os resultados são compatíveis com a resistência conhecida do fenômeno analisado a ser interpretado intencionalmente. Além disso, entender de modo mais preciso a maneira que a interpretação intencional falha também é de grande valia para a compreensão teórica de nossas atribuições de estados mentais.

Finalmente, terminei o capítulo analisando as possíveis consequências dos resultados encontrados no estudo de Rose *et al.* (2014) sobre a visão da PSC com relação a delírios. Cheguei à conclusão de que os dados encontrados pelo estudo não são suficientes para identificar as divergências com a TSI. Além disso, de modo ainda mais relevante, os dados encontrados ainda não foram suficientes para uma análise adequada da viabilidade da aplicação da postura intencional, pois ainda demanda-se explicações e previsões completas e precisas pela PSC e, também, da averiguação dos seus poderes explanatórios e preditivos.

Concluo, assim, que não puderam ser descobertos comportamentos que a PSC explica e que não podem ser explicados pela TSI, corroborando a tese de Dennett. Contudo, para averiguar esta situação de uma maneira mais adequada, seria crucial entender com mais clareza quais são os entendimentos da PSC para estes casos, quais previsões comportamentais são feitas, para não dizer de quais os seus modelos teóricos utilizados para realizar esta previsão. A partir de então, poderemos avaliar de modo mais contundente a viabilidade da aplicação da postura intencional como conceitualizada pela TSI de Daniel Dennett.

5. Conclusão da dissertação

Esta dissertação teve dois objetivos primários, a contribuição para o debate sobre a natureza da crença e o auxílio na elucidação da teoria de Daniel Dennett sobre os conteúdos mentais. A ideia inicial para a produção desta dissertação foi a de que casos-limite de atribuição de crença devido à irracionalidade dos sujeitos formam um ambiente propício para a averiguação da viabilidade da TSI. A hipótese que guiou o seu desenvolvimento era de que a investigação sobre o modo como ela lida com tais casos produziria evidências a respeito do rigor e da robustez de sua teoria ou a fraqueza e os limites da mesma.

Estimei que a evidência de uma possível fragilidade teórica adviria da sua suposta incapacidade para lidar com casos de falha de racionalidade. Para ir a fundo nesta linha de pesquisa, no segundo capítulo investiguei os princípios que guiam a utilização da postura intencional por meio da análise das principais críticas direcionadas à TSI em virtude do seu particular critério de racionalidade para a atribuição de crenças e outros estados mentais.

Ao se analisar estas críticas ao critério de racionalidade da TSI, identifiquei dois grandes grupos de argumentos: aqueles que se baseiam na dissimilaridade entre a PSC e a TSI; e aqueles baseados na inconsistência da aplicação do critério de racionalidade.

Concluí que a primeira crítica, do argumento da dissimilaridade entre a PSC e a TSI, não se sustenta no modo como foi formulada. Seria necessário mostrar que a PSC é capaz de explicar ou prever comportamentos que não podem ser explicados ou previstos pela TSI, mas os exemplos propostos, quando analisados cuidadosamente, não sustentaram esta conclusão. Já o argumento da inconsistência do duplo critério de racionalidade, como vimos, apesar de também não se sustentar do modo como foi formulado, tem o seu mérito, pois sinaliza uma falta de sistematização e explicação do modo de utilização do critério de racionalidade e do seu funcionamento dentro da aplicação. Concluo, portanto, que Dennett se defende das críticas direcionadas a sua teoria, porém percebe-se uma sentida ausência de sistematização para o uso do seu critério de racionalidade. Com base nesta conclusão e nos esclarecimentos oferecidos pela análise das críticas à TSI, conclui o segundo capítulo com a formulação de dois princípios que norteiam o funcionamento do critério de racionalidade da TSI em sua aplicação, o princípio da conexão holística e o princípio da redução da indeterminação.

Segundo Dennett, a sua teoria prevalece ou fracassa de acordo com os resultados empíricos de sua utilização. Portanto, um devido teste de sua teoria é crucial para análise de sua viabilidade. Como a crítica da dissimilaridade entre a PSC e a TSI não se sustentou por causa

de seu exemplo e a crítica do duplo critério sinalizou a falta de um esclarecimento do exato uso do critério de racionalidade, busquei desenvolver uma formulação clara e precisa de um método de aplicação da TSI, que leve em conta as peculiaridades do critério de racionalidade de sua teoria e que possa ser utilizado para qualquer situação que demande uma interpretação intencional, para, então, aplicá-la a novos exemplos que sejam mais desafiadores para a teoria. Para tanto, no terceiro capítulo me propus a aplicar a TSI a duas situações: de irracionalidade cotidiana, que é comumente encontrada e mais facilmente manejada, e de irracionalidade excepcional, aquela que compromete a própria interpretação e a atribuição de estados mentais. Para o primeiro escolhi um famoso caso de falha de raciocínio em uma situação em que as pessoas consistentemente realizam a falácia da conjunção, e para o segundo eu escolhi um caso de delírio clássico na literatura do debate sobre crenças, o delírio de Capgras.

Desse modo, esta linha de investigação que visa revelar as limitações da teoria das crenças de Dennett analisando casos periféricos de atribuição de crença também poderia trazer evidências da robustez de sua formulação teórica se ela se mostrasse capaz de acomodar os melhores e mais satisfatórios modelos explicativos para o entendimento deste fenômeno de marcante irracionalidade e de difícil atribuição de crença.

Tendo em vista este fim, desenvolvi um método para a aplicação da TSI que resultou na sua divisão em duas partes, cada qual com uma função particular para o uso da racionalidade na interpretação. Com a criação do primeiro nível, o nível de atribuição, pude contemplar a função da racionalidade para a redução da indeterminação da interpretação. Já no segundo nível, o nível de previsão, é onde a racionalidade tem a sua função de sustentar a rede de crenças do sistema que fundamenta qualquer previsão ou explicação intencional. Retomando as críticas analisadas anteriormente, conclui que a formulação deste método a partir dos dois níveis de aplicação esclarece de modo satisfatório como a TSI pode lidar com casos de irracionalidade.

Em seguida, parti para a aplicação do método encontrado para avaliar se a TSI pode demonstrar, de fato, que lida de modo satisfatório com casos de irracionalidade. Esta investigação, no entanto, encontrou um grande obstáculo na obtenção de seus objetivos. Mesmo com a formulação do método e a sua aplicação para os casos propostos, ainda seriam necessários alguns passos para demonstrar o sucesso ou o fracasso da TSI. Ainda seria necessário obter os resultados de uma explicação e previsão a partir da PSC, verificar uma possível divergência dos resultados encontrados para a TSI e, em caso afirmativo, comparar o sucesso preditivo ou explanatório de cada teoria.

A maior dificuldade encontrada foi em obter explicações e previsões representativas da PSC para as categorias de casos analisados. Para lidar com esta dificuldade realizei uma análise

provisória. Para o caso de irracionalidade cotidiana, utilizei as explicações propostas pelos autores da pesquisa e, para o caso de delírio, utilizei uma das escassas pesquisas da percepção da PSC para casos de delírios.

Os resultados obtidos no primeiro caso, de falha de raciocínio, mostraram mais claramente como a TSI pode dar conta de abarcar explicações que apontem para um funcionamento cognitivo sub-ótimo (com as limitações que acompanham a utilização de explicações pela postura de projeto para competências da postura intencional). Eles mostraram, também, que a TSI foi capaz de replicar as explicações alternativas propostas pelos autores da pesquisa. Este é um feito relevante pois estas explicações, além de terem sido corroboradas pelos resultados da pesquisa (sinalizando um poder explicativo), surgiram, estimo, a partir da utilização da PSC.

Já os resultados obtidos pela análise do caso de delírio expressaram a dificuldade encontrada pela TSI para explicar ou prever o comportamento de pessoas nesta condição. No entanto, esta conclusão não é problemática para a TSI porque casos de delírio são famosos pela sua resistência a se acomodarem em modelo teóricos explicativos. Considero que ainda há algo de positivo nos resultados da análise feita, pois eles lançam luz sobre o motivo de casos de delírio não se encaixarem em nossas interpretações. Além disso, a posterior comparação com os resultados do estudo da percepção da PSC para o caso de delírio de Capgras proporcionou uma ideia do que seria uma análise comparativa com os resultados obtidos pela PSC. No entanto, a análise deixou a desejar por causa que o entendimento a respeito da percepção da PSC sobre casos de delírio ainda é muito incipiente para possibilitar uma análise mais aprofundada.

A ambiciosa proposta de avaliar a viabilidade da TSI é um projeto que não pôde ser decisivamente concluído nesta dissertação. Considero, no entanto, que é um projeto em andamento e, espero, que a presente dissertação tenha contribuído de maneira interessante para a melhor compreensão da TSI, da aplicação do seu critério de racionalidade e das variáveis envolvidas na interpretação de um caso periférico de atribuição de crenças, de modo que pesquisas futuras possam se apoiar nas ideias e métodos aqui desenvolvidos para chegar mais longe na compreensão dos temas envolvidos.

REFERÊNCIAS

- ARMSTRONG, D. M. **A materialist theory of the mind**. New York: Routledge & Kegan Paul, 1968.
- BERMUDEZ, J. Normativity and rationality in delusional psychiatric disorders. **Mind & Language**, 16(5): 493—457, 2001.
- BORTOLOTTI, L. Can we interpret irrational behavior? **Behavior and Philosophy** 32(2): 359–75, 2004.
- _____. Delusions and the background of rationality. **Mind & Language**, 20(2): 189–208, 2005.
- _____. **Delusions and Other Irrational Beliefs**. Oxford: Oxford University Press, 2010.
- _____. **Irrationality**. Hoboken, NJ: John Wiley & Sons, 2014.
- BROWN, H. I. Rationality. In Honderich, T. (ed.) **The Oxford Companion to Philosophy**. New York: Oxford University Press, 1995.
- CHERNIAK, C. "Rationality and the Structure of Memory," **Synthese**, 57, pp. 163-86, 1983.
- _____. **Minimal Rationality**. Cambridge, MA: The MIT Press/A Bradford Book, 1986.
- CHISHOLM, R. M. **Perceiving**. Ithaca: Cornell University Press, 1957.
- CHURCHLAND, P. M. Eliminative materialism and the propositional attitudes. **Journal of Philosophy**, 78: 67–90, 1981.
- DAVIDSON, D. (1984). **Inquiries into truth and interpretation**. Oxford: Clarendon, 1984.
- DAVIES, M., Coltheart, M., Langdon, R., & Breen, N. Monothematic delusions: Towards a two-factor account. **Philosophy, Psychiatry, & Psychology**, 8(2), 133-158, 2001.
- DENNETT, D. C. Intentional systems. **The Journal of Philosophy**, 68(4): 87–106, 1971.
- _____. Mechanism and responsibility. In T. Honderich, ed. **Essays on Freedom of Action**. London: Routledge & Kegan Paul, 1973.
- _____. Brain Writing and Mind Reading. In K. Gunderson (ed.), **Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science**, Vol. VII. Minneapolis: University of Minnesota Press, pp. 403–416, 1975.
- _____. Conditions of personhood. in A. Rorty, (ed.) **The identities of persons**, 175-98, 1976.
- _____. **Brainstorms**. Cambridge, MA: MIT Press, 1978a.
- _____. Beliefs about beliefs. **Behavioral and Brain sciences**, 1(4), 568-570, 1978b.
- _____. Why you can't make a computer that feels pain. **Synthese**, 38(3), 415-456, 1978c.

- _____. Making sense of ourselves. **Philosophical Topics**, 12(1), 63-81, 1981a.
- _____. Three Kinds of Intentional Psychology. In R. Healey (ed.), **Reduction, Time and Reality**. Cambridge: Cambridge University Press, pp. 37–60, 1981b.
- _____. True Believers: the Intentional Strategy and Why it Works, in A. F. Heath, ed. **Scientific Explanation**. Oxford University Press, 1981c.
- _____. Intentional systems in cognitive ethology: The “Panglossian paradigm” defended. **Behavioral and Brain Sciences**, 6(3), 343-355, 1983.
- _____. **The Intentional Stance**. MIT press, 1987.
- _____. Out of the Armchair and into the Field. **Poetics Today**, 9(1), 205-221, 1988.
- _____. Real patterns. **The journal of Philosophy**, 88(1), 27-51, 1991.
- _____. An interview with Dan Dennett. **Cogito**, 6(3), 115-125, 1992.
- _____. **Back from the drawing board**. In Dennett and his critics: Demystifying mind. Blackwell: Oxford, 1993a.
- _____. **Consciousness explained**. Penguin uk, 1993b.
- _____. Get real. **Philosophical Topics**, 22(1/2), 505-568, 1994a.
- _____. Self-Portrait, in S. Guttenplan, ed., **A Companion to the Philosophy of Mind**, Blackwell Press, Oxford, 1994, pp. 236-244, 1994b.
- _____. **Darwin's dangerous idea**. *The Sciences*, 35(3), 34-40, 1995.
- _____. With a little help from my friends. **Dennett’s philosophy: A comprehensive assessment**, 327-388, 2000.
- _____. Intentional systems theory. In S. Walter & A. Beckermann (eds.), **The Oxford Handbook of Philosophy of Mind**. Oxford: Oxford University Press, pp. 339–350, 2009.
- _____. Not Just a Fine Trip Down Memory Lane: Comments on the Essays on Content and Consciousness. In C. Muñoz-Suárez & F. De Brigard (eds.), **Content and Consciousness Revisited**. New York: Springer, pp. 199–220, 2015.
- DRETSKE, F. Belief. In Honderich, T. (ed.) **The Oxford Companion to Philosophy**. New York: Oxford University Press, 1995.
- DUB, R. The Rationality Assumption. In C. Muñoz-Suárez & F. De Brigard (eds.), **Content and Consciousness Revisited**. New York: Springer, pp. 93–110, 2015.
- EGAN, A. Imagination, delusion, and self-deception. In T. Bayne & J. Fernández (eds.), **Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation**. Hove: Psychology Press, pp. 263–280.

- ELLIS, H. D.; YOUNG, A. W.; Quayle, A. H.; & De Pauw, K. W. Reduced autonomic responses to faces in Capgras delusion. **Proceedings of the Royal Society of London B: Biological Sciences**, 264(1384), 1085-1092, 1997.
- FODOR, J. A. Why paramecia don't have mental representations. **Midwest studies in philosophy**, 10(1), 3-23, 1987.
- FODOR, J.; LEPORE, E. Is Intentional Ascription Intrinsically Normative? **Dennett and his critics**, 70-82, 1993.
- FOLEY, R. **The theory of epistemic rationality** (pp. 214-15). Cambridge, MA: Harvard University Press, 1987.
- GENDLER, T.S. Alief and Belief. *Journal of Philosophy* 105: 634–663. Frankish, K. (2009). Partial belief and flat-out belief. In F. Huber & C. Schmidt-Petri (eds.), **Degrees of Belief**. Dordrecht: Springer, pp. 75–93, 2008.
- GERRANS, P. Delusions as performance failures. **Cognitive Neuropsychiatry** 6(3): 161-173, 2001a.
- KAHNEMAN, D. A perspective on judgment and choice: mapping bounded rationality. **American psychologist**, 58(9), 697, 2003.
- LEWIS, D. Psychophysical and theoretical identifications, **Australasian Journal of Philosophy**, 50: 249–258, 1972.
- MCKAY, R. T.; DENNETT, D. C. Our evolving beliefs about evolved misbelief. **Behavioral and Brain Sciences**, 32(6), 541-561, 2009.
- NICHOLS, S.; STICH, S. P. **Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds**. Oxford: Oxford University Press, 2003.
- NISBETT, R.E.; ROSS, L. **Human Inference: Strategies and Shortcomings in Social Judgement**. Englewood Cliffs, NJ: Prentice–Hall, 1980.
- PREMACK, D.; WOODRUFF, G. Does the chimpanzee have a theory of mind? **Behavioral and brain sciences**, 1(4), 515-526, 1978.
- PUTNAM, H. **Mind, language, and reality**. London: Cambridge University Press, 1975.
- QUINE, W. v. O. **Word and Object**. Cambridge, MA: MIT Press, 1960.
- ROSE, D.; BUCKWALTER, W.; TURRI, J. When words speak louder than actions: Delusion, belief, and the power of assertion. **Australasian Journal of Philosophy**, 92(4), 683-700, 2014.
- ROSE, D. et al. 'Behavioral Circumscription and the Folk Psychology of Belief: A Study in Ethno-Mentalizing', **Thought: A Journal of Philosophy**, 6 (3), 193-203, 2017.
- RYLE, G. **The concept of mind**. New York: Barnes & Noble, 1949.
- SCHWITZGEBEL, E. A phenomenal, dispositional account of belief. **Noûs**, 36(2): 249–275, 2002.

_____. Mad belief? **Neuroethics** 5(1): 13–17, 2012.

_____. **Belief**. Disponível em: <<http://plato.stanford.edu/entries/belief/#1.2>>. Acesso 01/10/2018, 2018.

STEIN, E. **Without good reason: The rationality debate in philosophy and cognitive science**. Clarendon Press, 1996.

STICH, S. P. Dennett on intentional systems. **Philosophical Topics**, 12(1), 39-62, 1981.

TVERSKY, A.; KAHNEMAN, D. Availability: A heuristic for judging frequency and probability. **Cognitive psychology**, 5(2), 207-232, 1973.

_____. Judgment under uncertainty: heuristics and biases. **Science** 185(4157): 1124–1131, 1974.

_____. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4), 293. Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. **American psychologist**, 58(9), 697, 1983.

VAN GELDER, T. Compositionality: A connectionist variation on a classical theme, **Cognitive Science** 14: 355–384, 1990.

WEBB, S. Witnessed behavior and Dennett's intentional stance. **Philosophical Topics** 22(1–2): 457–470, 1994.

UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS HUMANAS

PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA
ÁREA DE CONCENTRAÇÃO: **FILOSOFIA**

Por decisão do Colegiado do Programa de Pós-Graduação em Filosofia da UFPR, o aluno deverá atender as solicitações da banca, quando houver, e anexar este ao final da dissertação/tese como versão definitiva aprovada pelo orientador, que neste momento estará representando a Banca Examinadora.

Curitiba:

Prof. Dr. José Eduardo Freitas Porcher. Assinatura:

A handwritten signature in black ink, reading "José Eduardo Freitas Porcher". The signature is written in a cursive style and is positioned above a horizontal line.