UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE FILOSOFIA, CENTRO DE CENTRO DE FILOSOFIA E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

Matheus de Lima Rui

Em que um bayesiano acredita?: Uma investigação sobre a racionalidade de estados doxásticos qualitativos e quantitativos

Florianópolis-SC
2023

Matheus de Lima Rui

Em que um bayesiano acredita?: Uma investigação sobre a racionalidade de estados doxásticos qualitativos e quantitativos

Tese submetida ao Programa de Pós-Graduação em Filosofia da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Lógica e Epistemologia.
Orientador: Alexandre Meyer Luz, Dr.
Coorientador: Jonas R. Becker Arenhart, Dr.

Florianópolis-SC
2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Rui, Matheus de Lima
Em que um bayesiano acredita? : Uma investigação sobre a racionalidade de estados doxásticos qualitativos e quantitativos / Matheus de Lima Rui ; orientador, Alexandre Meyer Luz, coorientador, Jonas R. Becker Arenhart, 2023.
235 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Filosofia e Ciências Humanas, Programa de Pós-Graduação em Filosofia, Florianópolis, 2023.

Inclui referências.

1. Filosofia. 2. Epistemologia. 3. Lógica. 4. Probabilidade. 5. Racionalidade. I. Luz, Alexandre Meyer. II. Arenhart, Jonas R. Becker. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Filosofia. IV. Título.

Matheus de Lima Rui

Em que um bayesiano acredita?: Uma investigação sobre a racionalidade de estados doxásticos qualitativos e quantitativos

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Dr. Luis Rosa
University of Cologne

Prof. Dr. Cezar Mortari
Universidade Federal de Santa Catarina

Prof. Dr. Danilo Fraga Dantas
Universidade Federal da Paraíba

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Lógica e Epistemologia.

ICP-Edu
Documento assinado digitalmente
Jerzy Andre Brzozowski
Data: 14/07/2023 15:28:40-0300
CPF: ***.516.469-**
Verifique as assinaturas em https://v.ufsc.br

Coordenação do Programa de
Pós-Graduação

GOV.BR
Documento assinado digitalmente
ALEXANDRE MEYER LUZ
Data: 14/07/2023 14:30:10-0300
CPF: ***.644.799-**
Verifique as assinaturas em https://v.ufsc.br

Alexandre Meyer Luz, Dr.
Orientador

Florianópolis-SC, 2023.

Aos meus pais, Lúcia e Edson.

AGRADECIMENTOS

Antes de qualquer coisa, preciso reconhecer aqui que esses cinco anos de trabalho, incluindo uma pandemia no meio desse tempo, não foram fáceis, e que se esse trabalho teve um desfecho satisfatório foi fundamentalmente graças a todos que serão mencionados aqui, direta ou indiretamente.

Em primeiro lugar, preciso agradecer, como de costume, a todos familiares que me apoiaram nessa longa jornada longe do lugar onde nasci e cresci, fundamentalmente aos meus pais Edson Carlos e minha mãe Maria Lúcia, pelo apoio incondicional durante todo esse tempo. Ainda, devo agradecer especialmente a pessoa que passou esses cinco anos ao meu lado, nos momentos fáceis e difíceis, alegres e tristes, pelas discussões teóricas sérias e as nem tão sérias assim, e por aguentar minha incontrolável prática de criar trocadilhos ruins nos momentos mais importunos, minha companheira Luana Nyland. Agradeço por estarem presentes em mais uma fase importante da minha vida.

Em segundo lugar, preciso agradecer meus colegas que colaboraram para enriquecer vários dos tópicos que estão presentes nesta Tese. O fato de boa parte do doutorado ter acontecido remotamente, fez com que muitas oportunidades de interação fossem perdidas, e as mais marcantes pra mim foram, em sua maioria, interações remotas. Sob o risco de cometer alguma injustiça, vou mencionar aqui apenas algumas pessoas que contribuíram em períodos específicos durante a produção do presente material, mesmo que eles não tenham consciência disso: meu colega Valdenor Brito, por ser meu mentor inicial, desde o processo seletivo, até os primeiros passos dentro da UFSC; minha amiga e professora de inglês Celina Brod, pelos altos papos filosóficos durante nossas aulas, e por sempre me incentivar com os meus objetivos; meu amigo bayesiano André Neiva, que foi uma das únicas referências aqui no Brasil que eu tinha para conversar, e participar em conjunto de grupos de estudo, sobre epistemologia bayesiana; e também meu parceiro economista-matemático-filósofo Mateus Noriller pela inacreditável capacidade de achar legal alguns dos tópicos presentes nesta Tese.

Finalmente, preciso agradecer alguns dos professores que foram importantes nessa jornada. Primeiramente, gostaria de agradecer o PPGFIL-UFSC pelo apoio nesses anos. Primeiramente na pessoa do Prof. Dr. Ivan da Cunha, ex-coordenador do programa, por sempre me ajudar prontamente no que eu precisava. Também é essencial agradecer a gestão atual, Prof. Dr. Vilmar Debona e Prof. Dr. Jerzy Brzozowski, fundamentalmente por me auxiliarem no processo de concessão do auxílio viagem para participar e apresentar em um dos maiores eventos de lógica do mundo, o UNILOG, no ano de 2022, na ilha de Creta, na Grécia, que foi uma das experiências mais incríveis da minha vida. Dando sequência, preciso agradecer meu orientador Prof. Dr. Alexandre Meyer Luz por ter cultivado em mim a ideia do que é uma “atitude filosófica apropriada”, e por ter me encorajado e me apoiado na decisão de mudar o tema central da Tese, ainda no primeiro ano da pesquisa, o que foi a melhor decisão que eu poderia ter tomado naquele momento. Também devo agradecer meu coorientador Prof. Dr. Jonas Becker Arenhart por ter aceito o convite, no meio do caminho, para me ajudar nessa empreitada, sem ele me dando suporte eu não teria coragem de me aventurar por algumas temáticas que foram exploradas aqui. E finalmente, agradeço a imensa contribuição dos professores Luis Rosa e Cezar Mortari na leitura e avaliação do meu trabalho de qualificação, é quase desnecessário mencionar o privilégio que eu tenho em ser avaliado por teóricos desse calibre.

Agradeço à CAPES pelo suporte financeiro que me possibilitou a dedicação exclusiva ao trabalho de pesquisa aqui apresentado.

"At least some of my present philosophical beliefs will turn out to be incorrect"
(MAKINSON, D. C., 1965)

RESUMO

Na filosofia, e mais precisamente na Epistemologia, a forma dominante de se referir ao conceito de “crença” é tratando-o como um estado doxástico qualitativo: dada uma informação, o agente acredita em sua verdade, ou suspende o juízo. Todavia, também é possível compreender “crença” como um estado doxástico quantitativo, contendo graduações: duas informações acreditadas (de modo qualitativo) podem possuir distintos graus de confiança. Refiro-me a esse último estado como “probabilidade subjetiva”. Atualmente, existe uma extensa literatura sobre como cada uma dessas noções doxásticas devem ser entendidas. O objetivo do presente trabalho é explorar esse debate com o foco na racionalidade epistêmica de tais estados. Crenças (qualitativas) e probabilidades subjetivas possuem distintos padrões de avaliação racional: enquanto o primeiro tipo se aproxima de um padrão lógico dedutivo, o segundo é avaliado pela teoria da probabilidade clássica. Vários movimentos teóricos tentaram unificar essas duas noções, ou explicar como elas se relacionam do ponto de vista de sua racionalidade, mas a grande maioria das propostas, até então, esbarra em casos bem famosos de paradoxos que explicitam a latente inconsistência nesse empreendimento. A partir de uma revisão histórica e sistemática da literatura sobre tal desafio, este trabalho aponta as principais dificuldades e os caminhos mais promissores a se trilhar em direção a uma resposta satisfatória à questão levantada. Dentre as soluções possíveis, duas delas ganham maior atenção devido à relevância da tese epistemológica subjacente: a primeira aceita que conjuntos de crenças podem admitir inconsistências; já a segunda vai assumir que crenças são fortemente dependentes do contexto. Apresentadas as razões para a aceitação da segunda suposição, desenvolvo as consequências de uma tese que conecte crença e probabilidade subjetiva assumindo que a racionalidade de um conjunto de crenças é dependente do contexto. Por fim, exploro como o debate em torna da racionalidade de crenças e probabilidades subjetivas impacta nosso entendimento sobre teorias dinâmicas para o raciocínio, especificamente naqueles casos onde uma nova informação é adquirida.

Palavras-chave: crença; lógica; probabilidade; racionalidade.

ABSTRACT

In philosophy, and more precisely in epistemology, the mainstream way of discussing the concept of belief is to refer to it as a qualitative doxastic state: when presented with a piece of information, an agent either believes in its truth or suspends judgment. However, it is also possible to understand the concept of “belief” as a quantitative doxastic state, which can be graduated. Two distinct propositions that are believed (in a qualitative sense) can have different degrees of confidence. The latter is referred to as “subjective probability” (or “credence”). There is currently extensive literature on how each of these doxastic notions should be understood. The purpose of this work is to explore this issue with a focus on the epistemic rationality of these notions. Qualitative beliefs and subjective probability have distinct patterns of rational evaluation. While the former is closely associated with a logical-deductive pattern, the latter is evaluated by classical probability theory. Several theoretical movements have attempted to unify the two concepts or explain how they relate to each other in their rational point of view. However, most of them clash with well-known paradoxical cases that reveal the plain inconsistency underlying this type of undertaking. Through a historical and systematic review of relevant literature on the subject, this work draws attention to the main issues and most promising paths toward an acceptable answer to this question. Among the possible solutions, two of them receive more attention due to the relevance of their underlying epistemological theses. The first accepts that belief sets can admit inconsistency, while the second contends that beliefs are strongly context-sensitive. Given the reasons for accepting the latter, this work explores the consequences of a thesis that bridges belief and subjective probability, assuming that the rationality of a belief set is context-dependent. Finally, this dissertation examines how the debate over the rationality of qualitative and quantitative doxastic states affects our understanding of dynamic theories of reasoning, particularly in cases where new information is received.

Keywords: belief; logic; probability; rationality.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

INTRODUÇÃO

Considere o seguinte caso hipotético: Eu e minha esposa estamos de carro a caminho de casa. É fim de tarde de uma sexta-feira, e planejamos parar no banco para realizar um depósito. Entretanto, ao passar em frente, notamos uma grande fila dentro do banco, como é comum de acontecer sexta-feira à tarde. Desse modo, sugiro que voltemos no dia seguinte, sábado de manhã, para realizar o depósito. Minha esposa questiona, "Muitos bancos não abrem aos sábados, você *acredita* que estará aberto?". Enigmático como sempre, eu respondo, "Dado meu *alto grau de confiança*, eu estaria disposto a apostar 9 reais para cada 1 real que o banco estará aberto amanhã!". Nesse momento, a única coisa certa era a raiva que minha esposa sentiu ao ouvir minha resposta. Mas, teria eu, assim, respondido a pergunta de minha esposa? Alguns filósofos diriam que sim: minha alta confiança é suficiente para minha esposa acreditar que o banco estará aberto amanhã e ir direto para casa ao invés de passar um bom tempo na fila. Outros diriam que não: minha alta confiança não é suficiente para minha esposa tomar como verdadeira a afirmação de que o banco estará aberto amanhã. Outros ainda diriam que a *racionalidade* de acreditar ou não que o banco estará aberto no dia seguinte é algo que depende do contexto em questão e dos riscos envolvidos.

O caso apresentado é uma adaptação do famoso exemplo de Keith DeRose, em *Contextualism and Knowledge Attributions* (1992). Diferentemente do exemplo original, nosso foco não está em *atribuições de conhecimento*, mas sim, na *racionalidade de estados doxásticos*. Como vários elementos aqui merecem a devida atenção, deixe-me começar, primeiramente, introduzindo alguns dos conceitos relevantes.

Na filosofia, especialmente na epistemologia, quando dizemos que alguém acredita em algo, tradicionalmente atribuímos a ela um estado doxástico[1] tripartite. Isso quer dizer que trata-se de um estado com três possíveis *atitudes doxásticas*: o agente acredita, ou desacredita, ou suspende o juízo[2]. Quando os epistemólogos dizem que conhecimento requer "crença verdadeira justificada" eles estão se referindo a esse tipo específico de atitude doxástica. Contudo, não é incomum o fato de que, algumas vezes, nos referimos à crença em termos graduais: minha crença de que o Sol irá nascer amanhã é mais forte do que minha crença de que irá chover amanhã. Diferentemente

[1] Um estado doxástico é o estado mental comumente descrito pelo verbo "crer". Podemos assumir momentaneamente (por razões heurísticas) que tal estado toma como objeto sentenças descritivas, como por exemplo, "João *acredita que* faz sol agora em Florianópolis". Crenças também são consideradas "atitudes proposicionais", pois envolvem um tipo de atitude que um agente possui com relação a um pensamento (ou conteúdo), onde este pode ser tomado como o sentido de uma sentença descritiva. Detalhes mais finos sobre tais conceitos psicológicos fogem ao escopo deste trabalho, e, ao longo do texto, *atitudes* e *estados* serão geralmente tomados como referindo-se ao mesmo fenômeno doxástico.

[2] Ao longo do texto voltaremos com mais detalhes a esse tema, e veremos as razões para muitas vezes "crença" ser tratada como um estado com apenas duas alternativas: acredita ou não acredita.

do primeiro caso, essa última atribuição de “crença” é gradual, ela varia em uma escala com mais de três níveis, de um modo que a noção tradicional de crença é incapaz de representar. No primeiro caso temos um exemplo de estado doxástico *qualitativo*, e no segundo caso, um estado doxástico *quantitativo*.

De um modo um pouco mais intuitivo, quando, no exemplo anterior, minha esposa perguntou se eu *acreditava* que o banco estaria aberto no sábado, assumimos que ela queria saber se eu *aceitava como verdadeira* a afirmação de que o banco estaria aberto no dia seguinte. Na literatura em epistemologia, esse estado doxástico é conhecido de diversos modos, como, por exemplo, “aceitação”, “crença binária”, “crença cheia”, “crença qualitativa”, ou mesmo “crença *simpliciter*” (doravante, quando o termo “crença” aparecer sem nenhuma qualificação, estamos falando de sua versão qualitativa). Mas quando eu respondo expressando meu grau de confiança naquela afirmação, estou expressando uma atitude doxástica aparentemente distinta. Na literatura recente, teóricos se referem a tal estado de várias maneiras, como “graus de crença”, “crença gradual”, “crença parcial”, “crença quantitativa”, “probabilidade subjetiva”, ou também, pelo amplamente compartilhado termo anglo-saxão, *credence* (devido à falta de uma boa tradução do termo *credence* para o nosso idioma, daqui em diante, vou optar por “probabilidade subjetiva” para me referir à versão quantitativa). Ambos os casos são estados doxásticos capazes de expressar atitudes distintas, mais ou menos refinadas, tal que ambos dizem respeito a modos do agente emitir uma *opinião* sobre algo (vem daí o termo “doxástico”, do grego *doxa*, que traduzimos como “crença”, ou “opinião”).

Contudo, é importante dizer que esse não é, propriamente dito, um trabalho sobre “crença”, ou qualquer estado mental. A primeira advertência ao leitor sobre o objeto de estudo do presente trabalho é que não se trata de uma investigação sobre a *natureza* de um estado mental. Quem deve por sua própria função estudar a natureza dos estados mentais é a ciência cognitiva, uma vez que esse é um trabalho empírico, não próprio ao filósofo. Mas se esta investigação possui como objeto central isso que estamos chamando de crença (como sugere o título do presente trabalho), o que um filósofo pode dizer sobre o tema? Aqui aparece, acredito eu, nosso real objeto de investigação: o conceito de *racionalidade*. Voltando ao nosso exemplo inicial, podemos dizer que o que determinaria se eu, de fato, respondi ou não a pergunta de minha esposa é um *critério de racionalidade* para avaliar a atitude doxástica em questão. Portanto, meu objeto alvo será a racionalidade dos estados doxásticos, não sua “natureza”. Como será apresentado, tais estados não precisam, necessariamente, ser entendidos como entidades psicológicas, mas meras *idealizações* racionais. Pretendo deixar isso mais claro à frente, ainda na presente introdução.

Antes de qualquer coisa, é importante mencionar que estados doxástico dizem respeito ao modo como captamos o mundo e emitimos opinião sobre isso. Se nosso objetivo for captar corretamente os fatos no mundo podemos dizer que estamos interessados em estados *epistêmicos*. Nesse sentido, estados epistêmicos não são estados práticos, como, por exemplo, desejo, intenção, etc. Por consequência, o conceito de racionalidade em questão será teórico-epistêmico, e não prático. Assim, podemos partir do pressuposto de que existe um critério[3] de avaliação racional tanto para crenças quanto para probabilidades subjetivas: um critério genuinamente teórico-epistêmico. Esse tema aparecerá novamente e receberá o devido destaque ao longo de grande parte do presente trabalho.

Em resumo, encontramos na literatura atual basicamente dois modelos canônicos que fornecem padrões para a racionalidade da crença e probabilidade subjetiva: respectivamente, a Lógica Doxástica[4] e a Teoria da Probabilidade. Desse modo, sem muita surpresa, os modelos mais utilizados para a racionalidade da crença são modelos dedutivos. Por outro lado, a noção de probabilidade subjetiva é entendida, como o próprio nome diz, como uma noção probabilística, a qual desempenha tradicionalmente um papel relevante nas discussões sobre suporte indutivo, em contraste aos modelos dedutivos[5]. Na Seção 1.2, veremos como isso ocorre para uma noção de racionalidade específica. É a partir desse pano de fundo de suposições que nosso problema relevante se desenvolve.

A primeira pergunta a ser feita: qual a relação entre uma crença racional e uma probabilidade subjetiva racional? Alguém pode dizer que não há qualquer relação, que uma dessas atitudes é simplesmente resultado de um uso equivocado da nossa linguagem, ou também pode dizer que as duas existem independentemente uma da outra, e não há qualquer relação racional entre elas (Falaremos mais sobre esse tipo de proposta na Seção 1.3). Respectivamente, tais suposições fariam com que, em nosso exemplo inicial, ao ser questionado sobre minha crença e ter respondido expressando minha probabilidade subjetiva para o caso, temos que: ou a pergunta de minha esposa não faz sentido, ou minha resposta não faz sentido, ou as duas coisas fazem sentido, mas eu não respondi a pergunta dela em qualquer sentido relevante, uma vez que são dois estados totalmente independentes. No Capítulo 1, pretendo mostrar por que essas não são opções teórica atrativas.

Seguindo o espírito do exemplo inicial, é possível pensar na seguinte resposta: crença e probabilidade subjetiva se relacionam de modo que, quando possuo alta confiança, isto é, alta probabilidade subjetiva de que algo é o caso, posso afirmar

[3] Vamos deixar aqui essa afirmação propositalmente ambígua entre existe *ao menos um* ou *apenas um*. Isso será tratado com maior profundidade em Seção 1.2.3

[4] Entendo por Lógica Doxástica a parte da Lógica Epistêmica que diz respeito a atribuições de crença. Como veremos no Capítulo 4, a teoria AGM de revisão de crença também é uma opção compatível com o que será apresentado aqui para modelar a racionalidade da crença.

[5] Enquanto em modelos dedutivos a verdade das premissas garante, por meio de uma inferência válida, a verdade da conclusão, modelos de inferência probabilística podem garantir, no máximo, um maior grau de suporte para a verdade da conclusão.

(estou racionalmente autorizado) que tal e tal coisa é o caso. Segundo essa vertente, eu respondi a pergunta de minha esposa e foi uma resposta racionalmente apropriada, uma vez que minha probabilidade subjetiva era equivalente a 90%. Vamos chamar tal tipo relação subjacente à minha resposta de "Princípio Ponte": uma regra que conecta a racionalidade da crença com a racionalidade da probabilidade subjetiva. Como veremos, a tese de que crença racional equivale a alta probabilidade subjetiva é bastante intuitiva e aceita por uma parte da comunidade epistêmica. Contudo, ao menos em um primeiro momento, ela sofre com duras objeções.

Grosseiramente, ao assumir a tese do parágrafo anterior, há uma grande chance do agente acreditar, por fim, em um conjunto contraditório de afirmações. Isso aparece na literatura a partir dos famosos paradoxos do prefácio e da loteria, que serão propriamente apresentados na Seção 1.3. Assim, o objetivo primário da presente tese será investigar e apresentar quais são os possíveis princípios que conectam a racionalidade da crença com a racionalidade da probabilidade subjetiva, na forma de um Princípio Ponte. Veremos que cada proposta dá origem a desafios particulares. Na busca por um Princípio Ponte apropriado, veremos que as possíveis soluções tocam em teses historicamente relevantes para a epistemologia, como, por exemplo, a dependência da crença ao contexto, do fechamento dedutivo para crença, e da tese da crença como um estado epistemicamente estável. Por fim, também é importante mencionar como esses estados doxásticos se relacionam de um ponto de vista *dinâmico*, e não meramente *estático*. Tanto probabilidades subjetivas quanto crenças possuem seus respectivos modelos dinâmicos visando avaliar a racionalidade de um processo de atualização. Portanto, terminaremos nossa investigação vendo como regras de atualização racional se aplicam na construção de pontes entre as atitudes alvo.

Antes de passar para a apresentação dos capítulos, gostaria de fazer algumas considerações introdutórias sobre pressupostos e motivações de algumas das teses que estão sendo assumidas. A primeira motivação diz respeito ao papel "não-descritivo" do conceito de racionalidade: mesmo se fôssemos oniscientes sobre a funcionalidade da mente humana, isso nos diria muito pouco (ou nada) sobre o conceito de racionalidade que nos interessa! Tal afirmação pode fazer o leitor levantar as sobrancelhas, mas espero que seja apenas em um primeiro momento. Essa bifurcação entre a natureza da crença e sua racionalidade está ligada ao fato de que racionalidade é aqui assumida como uma noção aparentemente *normativa* (em um sentido amplo do termo): ela não descreve como os seres humanos raciocinam, mas, em alguma medida, ela avalia padrões ideais de inferência e consistência presentes no conjunto de estados mentais. Quando dizemos que a racionalidade "requer" algo de alguém, é a esse tipo de normatividade que estamos nos referindo. Esse padrão de correção, que chamamos de racionalidade, está presente no conjunto dos estados mentais de agentes humanos na medida em que ele soluciona alguns "desequilíbrios" dentro da teia de relações entre estados. Como afirma Peter Gärdenfors:

> Se o conjunto de crenças não é consistente, ou se a atribuição de probabilidade a um corpo de crenças não é coerente, então o indivíduo deve, se ele preenche os critérios da racionalidade, ajustar seus estados de crença até que eles atinjam um equilíbrio que satisfaça tal critério. Um estado racional de crença está em equilíbrio sob todas as forças do criticismo interno (GÄRDENFORS, 1988, p. 09, tradução nossa)[6].

Nesse sentido, o conceito de racionalidade assumido aqui não é nem mesmo uma exclusividade dos humanos! Mas antes que o leitor abandone a leitura por aqui, vou esclarecer melhor o que quero dizer com esta última afirmação.

Dentro do trabalho filosófico, é prática comum tomar estados epistêmicos racionais como idealizações de estados cognitivos *humanos*. Contudo, o conceito de racionalidade aqui pode ser estendido para *sistemas* idealmente racionais[7]. Desse modo, a racionalidade é um guia para solução de conflitos entre conteúdos proposicionais. Veremos como isso ocorre precisamente, e quais os desafios de cada proposta, ao longo de toda esta investigação.

Uma consequência interessante desse pressuposto é que, se essa noção de racionalidade fosse essencialmente conectada com a mente humana, seria necessário aceitar que o conceito de racionalidade atribuído a um sistema de inteligência artificial seria algo distinto, assim como a racionalidade de uma teoria, de grupo de agentes, de instituições, etc., seria algo diferente da racionalidade humana. Mas o uso comum do termo racionalidade parece se referir a algo que se aplica a uma multiplicidade de fenômenos, dos mais variados tipos. Não parece estar de acordo com nossas intuições sobre esse uso a ideia de que existam racionalidade*s* (embora possamos divergir bastante sobre qual o modelo mais apropriado). Ao dizer que um agente humano é racional, ou que uma teoria é racional, é razoável assumir que estou atribuindo uma mesma propriedade em ambos os casos. Desse modo, o critério de avaliação racional aplicado aos humanos assemelha-se (ou é algo bem próximo) ao critério que estaríamos dispostos a aplicar às máquinas inteligentes, e até mesmo a hipotéticos seres oniscientes, como anjos ou extraterrestres, que não possuam limitações cognitivas. Assim, racionalidade é assumida aqui uma propriedade de estados epistêmicos, não necessariamente uma avaliação de agentes humanos reais.

Uma vantagem desse pano de fundo teórico é que não precisamos assumir agentes ideais, como fazem boa parte dos epistemólogos formais que discutem esse tema[8]. Acredito que esse seja um avanço metodológico (mesmo que pequeno) do

[6] Ao longo de toda a presente Tese, as citações diretas de trabalhos publicados em língua estrangeira são traduções de autoria própria, salvo quando mencionado explicitamente o contrário.

[7] Uma visão bastante semelhante é compartilhada por Gärdenfors (1988).

[8] Ver por exemplo, Leitgeb (2017), Dorst (2017), Huber (2009) e Hawthorne (2009). Uma posição semelhante à que sustento aqui sobre o debate “idealização de estados *vs* a idealização de agentes” pode ser encontrada em Easwaran e Fitelson (2015). Em Titelbaum (2013, capítulo 4), também

presente trabalho em relação à literatura atual sobre o tema. Ao longo do texto, ao se deparar com exemplos de agentes ideais, é importante ter em mente que isso ocorre meramente por razões heurísticas, pois o que está sendo pressuposto é uma noção idealizada de racionalidade, não de agente. O que o agente humano faz com estados racionais, ou seja, a intrigante questão "por que ser racional?", não é uma questão que diz respeito propriamente à racionalidade, mas sim a outro campo de estudo que estuda "razões" normativas.

Avançando um pouco mais, como mencionado anteriormente, o conceito de racionalidade utilizado ao longo de todo o trabalho pode ser visto como um conceito normativo, no sentido em que racionalidade não é uma propriedade empírica presente em descrições do mundo. Resumidamente, é comum distinguirmos entre duas noções de normatividade: uma noção ampla que entende como normativo tudo aquilo que não é descritivo (ver, por exemplo, Skorupski (2007)); e outra, restrita, que entende normatividade como essencialmente atrelada a razões (ver, por exemplo, Broome (2013)). Não é difícil perceber que o segundo é um subconjunto do primeiro: tudo que diz respeito a razões é não-descritivo, mas nem tudo que é não-descritivo diz respeito a razões. Não por acaso, o conceito de racionalidade que assumimos se encaixa na última classificação: um conceito que é não-descritivo mas não diz respeito, necessariamente, a razões. Logo, para uma noção mais ampla de normatividade, a racionalidade pode ser entendida como normativa, mas para uma concepção mais restrita do termo, racionalidade não é normativa pois não acarreta, por si só, qualquer juízo prescritivo na forma de um "deve" [9].

Como veremos ao longo do primeiro capítulo (mais especificamente, na Seção 1.1), existem disputas teóricas sobre qual conceito de racionalidade é mais apropriado para dar conta do objeto de estudo em questão, e essa disputa envolve o modo como entendemos a relação entre racionalidade e normatividade. Nossa preferência será por uma noção de "racionalidade estrutural". Essa visão carrega consigo a ideia de que a racionalidade é fonte de "requisitos racionais"[10]. Tal aprofundamento é im-

encontramos uma excelente exposição desse debate, com uma posição bastante próxima a assumida aqui. Uma explicação mais detalhada sobre tal pressuposto metodológico e uma breve contribuição minha ao debate pode ser encontrada no Apêndice A.

9 Alguns autores utilizam o termo "avaliativo" em contraste com "normativo" (por exemplo, Easwaran e Fitelson (2015) retomam essa distinção do trabalho de Smith (2005)). Segundo essa visão, juízos normativos implicam em normas e prescrições, enquanto juízos avaliativos envolvem apenas conceitos como "bom", "correto", a partir de algum ideal de avaliação, e não constitui um guia para o agente. Contudo, acredito que tal distinção não seja suficiente. Por exemplo, segundo Broome (2013, cap. 2), a noção fundamental de normatividade está atrelada ao conceito de razões e ao "deve" (*ought*) prescritivo. Contudo, existe uma infinidade de "normas" que são "normativas" (no sentido de servirem de guia para o agente) mas que não possuem uma relação estrita com razões ou acarretem algum tipo de "deve". Por exemplo, existem normas constitutivas, regulativas, normas técnicas (Cf. Wright (1963)) que não são meramente avaliativas, mas que não fornecem, por si só, qualquer razão para seu cumprimento. Para meu objetivo, a noção menos rígida de "não-descritivo", como contrastando com "normativo", é mais abrangente e se encaixa bem com o presente propósito.

10 Essa tese tem como um de seus principais defensores o filósofo John Broome (2013, 1999).

portante na medida em que não precisaremos assumir um agente humano ideal para modelar uma noção idealizada de racionalidade. Qualquer tipo de conceito normativo (principalmente o "deve") que surja em decorrência de uma atribuição de racionalidade possuirá meramente finalidade heurística, dado que os conceitos de racionalidade e normatividade não precisam estar conectados.

Ainda, algum leitor pode se sentir desconfortável com as suposições anteriores e sustentar que uma investigação sobre racionalidade ideal não nos ajuda em nada na compreensão de fenômenos humanos, sobre como agentes epistêmicos reais funcionam, e como eles deveriam proceder em ambientes teóricos e práticos. Ou seja, alguém poderia alegar que o presente trabalho não teria qualquer propósito significativo na realidade. Apesar da mesma alegação poder ser direcionada para grande parte dos trabalhos em filosofia, acredito que ela não possui respaldo aqui, e espero deixar isso claro ao longo do texto. E para o leitor que ainda não estiver convencido, recomendo fortemente a leitura do Apêndice A, onde faço uma discussão mais aprofundada sobre esse debate metodológico.

Como veremos, meu interesse se concentra em alguns fenômenos específicos sobre a racionalidade de estados epistêmicos. Principalmente, pretendo apresentar como se conectam modelos quantitativos e qualitativos na representação formal de atitudes doxásticas. E, em muitos casos, isso será feito partindo de um ponto de vista de um agente ideal (embora, como foi dito acima, isso não precisa necessariamente ser o caso). Para isso, inicialmente serão utilizadas ferramentas matemáticas como Lógica Doxástica e Teoria da Probabilidade para avaliar, e formalizar, o conjunto doxástico do agente. Ao longo de todo o texto veremos como ferramentas formais possuem não somente um papel na avaliação de critérios de racionalidade, mas também são parte da modelagem formal do fenômeno em questão, isto é, da racionalidade dos estados epistêmicos alvo. Novamente, não se pode saltar dos resultados aqui apresentados como acarretando algo para agentes reais, nem mesmo para agentes ideais: primeiramente é preciso uma *interpretação* apropriada do modelo formal para ser aplicado em estados psicologicamente reais ou possíveis, para determinar a racionalidade desses estados[11]. Em segundo lugar, passar da racionalidade dos estados para a racionalidade dos agente (reais ou ideais) necessita de uma tese adicional sobre o que o agente deve fazer com a racionalidade, em outras palavras, se a racionalidade é normativa no sentido de nos proporcionar razões.

Para finalizar, considero a seguinte analogia esclarecedora: fazer filosofia da racionalidade consiste em retirar o fenômeno de seu estado de natureza e levá-lo ao laboratório. Neste caso, um laboratório de "filosofia formal". Ao longo do trajeto, pode ser que algumas coisas se percam no caminho. Pode ser que o fenômeno no ambiente controlado não funcione exatamente como em seu estado de natureza. Cabe

[11] No Apêndice A eu desenvolvo com a devida atenção tal afirmação.

então ao filósofo e ao epistemólogo dizer o que está acontecendo no ambiente controlado, e o que podemos transpor de um ambiente ao outro. Neste trabalho, o objeto de estudo será a racionalidade do sistema doxástico de um agente epistêmico, mais precisamente, dos estados qualitativos da crença, e dos estados quantitativos da probabilidade subjetiva, ambos analisados do ponto de vista de um ambiente controlado, ou podemos dizer, de um ambiente idealizado.

APRESENTAÇÃO DOS CAPÍTULOS E CONTRIBUIÇÕES AO TEMA

Como vimos, o ponto central da presente investigação gira em torno da busca de um Princípio Ponte para crença e probabilidade subjetiva a partir de uma noção de racionalidade específica. A linha investigativa do presente trabalho avançará sobre as possíveis respostas à seguinte questão: em que medida a resposta à minha esposa, no exemplo que inicia a introdução, foi realmente uma resposta racionalmente apropriada? No primeiro capítulo será apresentada com mais detalhes a noção de racionalidade estrutural e o conceito de “requisito racional” por trás dessa noção. Pretendo mostrar que existem motivações para essa escolha, e que olhar para o conceito de racionalidade a partir de requisitos racionais pode nos auxiliar a fundamentar mais precisamente as propostas que veremos nos capítulos seguintes.

Definido o conceito de “requisito racional”, podemos pensar nas exigências para a racionalidade da crença e da probabilidade subjetiva. Para isso, recorro à literatura em epistemologia formal sobre tais noções. Na Seção 1.2 será introduzido o pano de fundo formal com o qual irei trabalhar durante toda a tese. Principalmente, uma Lógica da crença (compatível com a Lógica Doxástica) irá guiar boa parte da investigação sobre a racionalidade da crença, enquanto a Teoria da Probabilidade será o guia da racionalidade da probabilidade subjetiva. Por fim, no final do primeiro capítulo será apresentado um dos principais desafios da presente tese: lidar com os paradoxos gerados pelas suposições anteriormente assumidas sobre a racionalidade dos estados doxásticos. A partir disso, será possível traçar algumas rotas investigativas a serem perseguidas nos capítulos seguintes. Ao apresentar os elementos essenciais de nosso pano de fundo formal, o presente trabalho também se propõe a oferecer uma introdução à epistemologia formal com foco no uso da lógica e da teoria da probabilidade visando modelar fenômenos epistêmicos. Tal contribuição pode ser reconhecida como uma das únicas introduções (até o presente momento, e até onde fui capaz verificar) sobre o tema em língua portuguesa, mas também oferece considerações ainda não exploradas na literatura global, fundamentalmente de língua inglesa, considerações essas que pretendo deixar explícitas ao longo do primeiro capítulo.

Dadas as considerações sobre os pressupostos teóricos e o nosso problema alvo, no segundo capítulo começo propriamente a trilhar alguns caminhos em busca de respostas. Em um primeiro momento, com um propósito histórico e elucidativo, será reconstruído o debate inicial sobre a relação entre crença e probabilidade subjetiva. Tal recorte histórico inicial é motivado, e acredito que justificado, pela inexistência atualmente de uma literatura que explora a origem das *Teorias da Utilidade Epistêmica* e sua relação com as primeiras tentativas de oferecer uma resposta sobre a relação entre a racionalidade da crença e da probabilidade subjetiva. Desse modo, veremos como o surgimento do debate em torno do tema estava ligado a preocupações relativas ao método científico, principalmente atreladas ao conflito entre modelos dedutivos e indutivos. Juntamente a esse movimento inicial surgiu também o que ficou conhecido por *Teoria da Decisão Cognitiva*, ou *Teoria da Decisão para Crença*. Segundo seus principais expoentes (HEMPEL, 1962; LEVI, 1967a), a relação entre crença e probabilidade subjetiva envolve elementos de utilidade epistêmica, como o conteúdo informacional de uma afirmação, e também, em alguns casos, elementos não-epistêmicos, como a inclinação ou aversão do agente ao risco. Ao longo do Capítulo 2, veremos como essa discussão se desdobrou contemporaneamente, e como uma Teoria da Decisão Cognitiva pode ser relevante ao oferecer ferramentas e novos *insights* aos desafios na busca de um Princípio Ponte adequado.

No Capítulo 3, pretendo aprofundar a discussão em torno de duas versões de Princípio Ponte que vêm recebendo um significativo destaque no debate recente: trata-se da “Tese Humeana da Crença” desenvolvida por Hannes Leitgeb (2017), e da rival “Teoria ProbaLógica” desenvolvida por Hanti Lin e Kevin Kelly (2012). Veremos como ambas visões estão comprometidas com uma noção de crença dependente do contexto e são capazes de fornecer um Princípio Ponte formalmente apropriado, mantendo a consistência no conjunto de crenças, frente aos paradoxos apresentados no primeiro capítulo. Também será explorado o debate em torno do elemento contextual da racionalidade da crença, e como isso está conectado com a suposição de que a crença é um estado “estável”. O capítulo se encerra com uma contribuição, parcialmente original, do que acredito ser a mais dura objeção à teoria de Leitgeb: ou a teoria será demasiadamente cética, ou, nas palavras de Schurz (2019), demasiadamente “mente fechada”. E por fim, o terceiro capítulo também oferece uma introdução inédita em nossa língua sobre tal discussão, e também uma das mais amplas e detalhadas sistematizações sobre o debate entre as teorias de Leitgeb e Lin & Kelly nos outros idiomas (até onde fui capaz de encontrar).

Como veremos, os capítulos 1,2 e 3 são restritos às normas “sincrônicas” (isto é, estáticas) da racionalidade da crença e da probabilidade subjetiva. O capítulo 4 adicionará o elemento “diacrônico” (isto é, dinâmico) no problema em questão. Veremos quais os desafios para um Princípio Ponte dinâmico, isto é, um princípio que leva em conta a atualização de crenças por nova evidência ao longo do tempo. As propostas investigadas serão desenvolvidas a partir das duas teorias do capítulo anterior, e como elas oferecem respostas rivais a esse desafio. Com isso, espero mostrar que um princípio ponte adequado que conecte a racionalidade da crença e da probabilidade subjetiva pode gerar algumas consequências interessantes para discussões, tanto em epistemologia tradicional, quanto no debate em epistemologia formal.

Por fim, encerro o Capítulo 4 apresentando uma breve (dis)solução para o conflito entre as abordagens rivais (a Tese Humeana da Crença e a Teoria Probalógica) que surge na tentativa de oferecer um modelo de Princípio Ponte diacrônico. Acredito que podemos olhar para tal conflito relevante de um ponto de vista metodológico, onde ambas as teorias, apesar de incompatíveis, oferecem modelos distintos que podem servir para distintos propósitos. Tal concepção está baseada, em grande parte, nas considerações metodológicas que estão contidas no Apêndice A, onde eu exploro a discussão sobre o uso de ferramentas formais no trabalho epistemológico, e ofereço uma posição teórica que é a assumida durante toda a presente Tese como um pano de fundo. Considero que tal visão metodológica seja a alternativa mais interessante aos problemas apresentados no referido apêndice e que seja capaz de nos orientar por vários problemas neste trabalho.

1 A RACIONALIDADE DE DUAS ATITUDES DOXÁSTICAS

Embora o termo "racionalidade" tenha suas raízes na literatura filosófica e científica, grande parte da população é capaz de compreender o sentido do termo quando ele aparece em frases e situações cotidianas, como, por exemplo, "Você está sendo muito irracional!", "Dado tudo que ocorreu, só me restou agir racionalmente!", "Eu não acredito nisso, essa me parece uma teoria muito irracional". Podemos dizer que o uso do termo "racionalidade" aparece tanto em contextos técnicos, como na filosofia e nas ciências, como na linguagem ordinária. Contudo, em nenhum desses locais encontramos uma uniformidade em seu uso. Como já mencionei brevemente na introdução, este é um trabalho sobre racionalidade epistêmica: sobre a racionalidade do modo como lidamos com as informações adquiridas sobre o mundo visando ter crenças verdadeiras. Ao longo deste capítulo, espero que isso fique mais claro.

Este capítulo será constituído por três grande partes. Na primeira, inicio explorando parte da literatura relevante sobre concepções de racionalidade. Veremos que pequenas distinções no modo como entendemos o conceito geram diferenças teóricas significativas. Pretendo convencer o leitor de que existe uma noção relevante de racionalidade que parece desempenhar uma função importante no debate sobre a racionalidade da crença e da probabilidade subjetiva. Isso será feito a partir da noção de racionalidade estrutural e do conceito de "requisitos racionais". Na segunda parte, apresento com mais detalhes uma possibilidade, até então não abordada pela literatura relevante, de enriquecer o debate sobre requisitos racionais a partir de elementos de epistemologia formal sobre a racionalidade da crença e da probabilidade subjetiva. Acredito que a literatura sobre requisitos racionais e a literatura em epistemologia formal podem se beneficiar mutuamente com tal aproximação. Além disso, boa parte do pano de fundo formal assumido ao longo do presente trabalho será apresentado aqui[1], o que torna o presente capítulo uma introdução ao tema para um público não especialista, ou não familiarizado com o uso de ferramentas formais em filosofia. Por fim, na terceira e última parte do capítulo, veremos os desafios que surgem ao aceitar todas as suposições feitas sobre a racionalidade dos estados doxásticos de crença e probabilidade subjetiva e, a partir daí, traçar um caminho para responder às principais questões levantadas durante o capítulo, e também na Introdução.

1.1 RACIONALIDADE ESTRUTURAL *VS.* RACIONALIDADE SUBSTANCIAL

Suponha que eu acredite na seguinte declaração: "chove agora em Florianópolis". Boa parte da comunidade filosófica, principalmente entre lógicos e epistemólogos, aceitaria facilmente a existência de uma norma de racionalidade que exija o seguinte: dado minha crença inicial, seria irracional acreditar também que *não* chove agora em

[1] O Capítulo 4 terá algumas exceções relevantes, que serão apresentadas no momento oportuno.

Florianópolis, ou qualquer outra afirmação que implique diretamente essa última. Nem tão consensual, no entanto, é a explicação que cada um oferece para explicar a fundamentação dessa norma da racionalidade. É em torno dessas divergências que a presente investigação se inicia.

Foram mencionadas anteriormente, na introdução, algumas nuances sobre o conceito de racionalidade. Neste tópico, exploro mais detalhadamente esse debate, agora em torno da disputa entre duas visões distintas sobre como o conceito de "racionalidade" deve ser entendido. Por um lado, existem aqueles que são chamados de *substancialistas* (que defende a tese da racionalidade substancial), que afirmam que ser racional é agir conforme as razões adequadas. Por outro lado, temos os *estruturalistas* (defensor da tese da racionalidade estrutural) afirmando que a racionalidade consiste na ampla coerência entre estados mentais, ou atitudes proposicionais[2]. Nas palavras de Tim Scanlon,

> Chamarei afirmações desse tipo de afirmações *estruturais* sobre a racionalidade, para distingui-las de afirmações *substantivas* sobre o que é uma razão para algo. Elas são estruturais pois são afirmações sobre relações entre atitudes do agente que devem sustentar-se na medida em que ele, ou ela, não é irracional, e o tipo de irracionalidade envolvida é uma questão de conflito entre tais atitudes (SCANLON, 2007, p. 85, grifo do autor).

Retomando a intuição extraída do nosso exemplo anterior, sobre o tempo em Florianópolis, o substancialista diria que uma norma de não-contradição se fundamenta no fato de que é impossível que um agente racional possua *razões* para acreditar em algo e em sua negação simultaneamente. Já o estruturalista diria que acreditar em contradições é irracional em função da *forma lógica* implícita no conteúdo da própria declaração. Veremos a seguir as implicações de cada uma dessas alegações.

Minha principal contribuição neste capítulo será mostrar que a escolha de uma noção específica de racionalidade dentro da disputa epistemológica relevante pode influenciar os desdobramentos teóricos do presente trabalho. Veremos como tal caminho teórico influencia o debate sobre a racionalidade da crença e da probabilidade subjetiva, como é possível estabelecer critérios de racionalidade estrutural a partir do conceito de "requisitos racionais", e quais as implicações disso. Com isso, pretendo desenhar um pano de fundo teórico com uma teoria da racionalidade que fundamente os desafios que aparecerão ao longo da presente Tese. Mirando esse alvo, portanto, se faz necessário entender um pouco melhor no que consiste esse conflito teórico entre a tese substancial e estrutural da racionalidade.

Agora de um modo mais preciso, a tese da racionalidade substancial sustenta que responder corretamente a razões é uma condição necessária e suficiente para a

2 A origem da distinção entre racionalidade estrutural e substancial aparece inicialmente nos trabalhos sobre racionalidade prática, principalmente com Scanlon (2007) e Hooker e Streumer (2004).

racionalidade[3]. Ou seja, segundo essa visão, existe uma equivalência entre ser racional e responder corretamente a razões. Encontramos alguns exemplos dessa posição na literatura epistemológica, com Robert Nozick afirmando que "Racionalidade envolve considerar razões favoráveis e contrárias" (NOZICK, 1993, p. 72), mas principalmente na literatura sobre racionalidade prática, com Joseph Raz afirmando que "Racionalidade é sobre a percepção do que são, ou são tomadas como, razões normativas, e a resposta a elas" (RAZ, 2005, p. 06). Mas como devemos entender aqui o que são "razões"?

O termo "Razões" desempenhará um papel técnico, e existe uma vasta literatura sobre suas possíveis interpretações. Apesar das discordâncias, podemos dizer que há certo consenso sobre algumas de suas atribuições. A primeira delas é sobre a distinção entre razões teóricas e razões práticas[4]: do lado prático, dizemos que crenças e desejos podem servir de razões para a formação de uma intenção (visando ação); do lado teórico o termo aparece geralmente conectado ao conceito de "evidência" para crença[5]. Por se tratar de um trabalho de epistemologia, trabalharei basicamente com questões do domínio teórico.

A segunda distinção importante é entre razões "motivacionais" e "normativas"[6]. Razões normativas são entendidas como razões "objetivas": ser uma razão normativa para a crença de que chove agora em Florianópolis implica na existência de um fato que conte como evidência para tal crença, independentemente da minha crença ou não nesse fato. Por outro lado, razões motivacionais são vistas como "subjetivas", ou seja, constituem uma explicação psicológica: ser uma razão motivacional para a crença de que chove agora em Florianópolis implica na mera existência de uma outra crença que conte como evidência para a minha primeira, não sendo necessário que essa crença esteja baseada em algum fato. Um exemplo deste último tipo seria um agente que acredita que chove agora em Florianópolis pois simplesmente sonhou com isso.

Retomando novamente o exemplo inicial, um substancialista diria que possuir um conjunto contraditório de crenças é irracional pois não há um conjunto de evidências que justifique as duas crenças ao mesmo tempo. Por exemplo, não existe um conjunto de evidências que justifique, no mesmo instante, tanto minha crença de que chove agora em Florianópolis, quanto de que não chove agora em Florianópolis. Desse modo, entendendo que evidências são razões epistêmicas, a visão substancial da racionalidade defende que uma crença racional é simplesmente uma crença formada por

[3] Tal posição pode ser encontrada implicitamente em grande parte da tradição filosófica, e explicitamente em alguns autores, como, por exemplo, Gibbard (1990) e Nozick (1993), e principalmente Raz (2005) e Kolodny (2005, 2007).

[4] Exemplos podem ser encontrados em Bratman (1987), Broome (2013) e Harman (1986).

[5] Por enquanto, o termo "evidência" possui aqui um sentido bem amplo: um fato que "conta a favor" para a verdade de uma determinada afirmação.

[6] Segundo Wedgwood (2017, p. 87), grande parte dos filósofos que têm discutido sobre "razões" distinguem entre razões motivacionais e razões normativas. Minha exposição aqui será baseada na distinção do autor.

razões apropriadas, sustentando assim uma equivalência entre razões e racionalidade. Podemos notar como isso aparece em um dos principais defensores atualmente da tese substancialista:

> O problema não é, como a ideia de requisitos formais de coerência sugere, que atitudes incoerentes estão em desacordo *umas com as outras*. Ao invés disso, é que quando atitudes são incoerentes segue-se que uma dessas atitudes está em desacordo com as *razões* para ela - o que ocorreria mesmo se elas não fossem parte de um conjunto incoerente (KOLODNY, 2007, p. 231, grifo do autor).

Por outro lado, o teórico estruturalista[7] questiona essa equivalência. Para ele, o campo das razões e o campo da racionalidade não precisam estar necessariamente juntos. Por exemplo, ainda que não exista qualquer evidência sobre a presença ou ausência de chuva em Florianópolis, a irracionalidade de um conjunto contraditório de crenças é determinada em função da própria estrutura lógica do conjunto. Basicamente, a tese da estruturalista é motivada pela suposição da insuficiência da tese substancialista.

A primeira alegação do estruturalista é que a versão normativa do conceito de razões é demasiadamente "objetiva" para ser parte essencial da definição de racionalidade. Tal visão seria incompatível com uma tese bem compartilhada entre a comunidade filosófica que discute racionalidade: a tese de que a racionalidade diz respeito exclusivamente ao conteúdo mental do agente. Sendo mais preciso:

Definição 1 (Tese da Sobreveniência da Racionalidade Sobre o Mental)**.** Nenhum elemento não-mental pode influenciar a racionalidade de um estado, ou de um conjunto de estados[8].

Essa tese sustenta basicamente que, se sua mente em determinada situação possui as mesmas propriedades do que em outra situação, então você é racional em uma situação no mesmo grau em que é racional na outra. Isso implica que nenhum ato não-mental está sujeito às exigências da racionalidade. Isto é, o que a racionalidade requer de determinado indivíduo é de ordem própria de sua mente. Para nós, seres humanos[9], racionalidade é algo que só podemos atribuir a estados mentais, e isso exige que eles estejam relacionados uns com os outros. Dizer isso é afirmar que a racionalidade requer que seus estados mentais sejam coerentes em aspectos particulares.

Uma vez que razões normativas para crer em algo sempre escapam, em algum nível, do agente, há muitas razões que não fazem parte do conjunto atual de estados

[7] Encontramos exemplos em Bratman (1987), Brunero (2010) e Broome (2013).

[8] Tal tese aparece explicitamente em Wedgwood (2002a) e Broome (2013, p. 151), embora seja implicitamente aceita por grande parte da literatura filosófica sobre racionalidade.

[9] Como foi dito na introdução, acredito que a noção de racionalidade assumida aqui é ampla o suficiente para ser aplicada a diversas coisas não-humanas, como máquinas, teorias, extra-terrestres, etc. Ao leitor interessado sobre o tema, recomendo a leitura do Apêndice A da presente Tese.

do respectivo agente. Em outras palavras, existe uma infinidade de evidências para acreditar em coisas que estão completamente fora de nosso alcance, em determinadas situações. Assim, se ser racional equivale a responder corretamente a razões, a racionalidade não pode ser uma propriedade que sobrevêm ao conjunto de estados do agente. Contudo, o filósofo substancialista pode aceitar que o termo “razões”, presente em sua definição de racionalidade como “responder corretamente a razões”, diz respeito a “crenças sobre razões”, ou seja, no que o agente acredita que é uma razão para crer. Em outras palavras, a racionalidade de uma crença irá depender do que o agente acredita que é uma evidência para crer naquilo. Essa seria uma versão da definição de razões motivacionais, e não necessariamente normativa, por sua vez, compatível com a tese da sobreveniência da racionalidade sobre o mental.

Para finalizar essa modesta sequência de objeções à tese substancialista, abordarei brevemente mais uma alegação sobre a insuficiência das razões na definição da racionalidade. Isso significa que, mesmo aceitando que crenças sobre razões são necessárias para a racionalidade, elas parecem não ser suficientes. A primeira alegação surge a partir de um caso de racionalidade prática. Suponha que um exótico terrorista lhe aponte uma arma e diga que irá tirar sua vida a menos que você intencione *p* e *não-p* (se isso é uma possibilidade psicológica real, há dúvidas, mas parece não haver problemas quanto a sua possibilidade lógica)[10]. Nesse caso, tudo indica que a prudência seria uma razão predominante sobre a razão para ser racional. Nesse caso, se é uma exigência da racionalidade não ter intenções contraditórias, é possível ser irracional mesmo respondendo corretamente a razões. Vejo que esse caso já seria um bom motivo para rejeitar a tese substancialista. Apesar de ser um exemplo prático, se pretendemos preservar uma continuidade entre racionalidade prática e teórica, então uma objeção no âmbito prático à tese substancialista já consiste em uma objeção sobre a racionalidade de tal tese.

Contudo, existem casos estritamente epistêmicos que podem ser tomados como contraexemplos à tese substancialista sobre a suficiência das razões. Como veremos no final do presente capítulo, mais especificamente na Seção 1.3, existem casos onde, ao responder corretamente a razões, somos conduzidos (ao menos aparentemente conduzidos) a um conjunto contraditório de crenças. Esses são paradoxos de consistência, conhecidos na literatura principalmente por meio dos paradoxos do prefácio e loteria. Uma discussão detalhada desses casos aparecerá mais à frente. Todavia, apesar de contraintuitivo, um teórico substancialista poderia alegar que não há nada de irracional em possuir um conjunto de crenças contraditórias, ou mesmo intenções contraditórias. Essa resposta possui detalhes com implicações bastante relevantes para o desenvolvimento do presente trabalho. Detalhes suficientes para merecer um

[10] Encontramos em Broome (2013, cap. 09) uma discussão mais ampla sobre a existência de intenções contraditórias em um mesmo agente.

tópico especial no próximo capítulo (Seção 2.3).

É claro que isso não esgota todo o debate sobre o tema. Contudo, espero ter apresentado algumas das motivações para se aceitar uma noção de racionalidade estrutural para racionalidade de estados doxásticos (o que será desenvolvido no próximo tópico). Desse modo, assume-se aqui que racionalidade não pode ser simplesmente responder corretamente a razões. É necessário tratar a racionalidade como uma fonte autônoma de exigências, independente das razões existentes para segui-las. Daqui em diante, a racionalidade será abordada como *fonte* de um conjunto de normas de consistência visando a harmonia entre os estados (ou o "equilíbrio, tal como abordado na introdução).

1.1.1 O que são requisitos racionais estruturais?

De um modo ainda não totalmente preciso, podemos dizer que a racionalidade requer que um sistema de crença não possua inconsistências. Um conceito utilizado para representar esse tipo de norma da racionalidade estrutural é conhecido na literatura por "requisitos racionais", principalmente devido ao trabalho pioneiro de John Broome (1999, 2013). Daqui em diante, quando se falar em requisitos racionais, estou me referindo a um termo técnico: eles especificam conflitos em nossas atitudes, os quais devemos resolver (ou evitar) se pretendemos ser racionais. Pontualmente, um requisito racional de não-contradição pode ser apresentado do seguinte modo:

Definição 2 (Não-Contradição)**.** É exigido de um agente racional que, se ele acredita que *p*, então ele não acredita que *não-p*.

É importante notar um aspecto desse requisito: se um agente já acredita em *p*, não é necessariamente o caso que ele "deva" desacreditar *não-p*, nem que ele tenha uma razão para isso. Por exemplo, imagine que você não deva acreditar em *p*, mas, ainda assim, você o faz (por exemplo, pode ser que a crença em *p* seja fruto de *wishful thinking*). Logo, não é o caso que você deva desacreditar *não-p*, ou que tenha uma razão para isso. Pode ser simplesmente que você não possua qualquer evidência sobre a verdade ou falsidade de *p*, e assim, deva suspender o juízo. Tal como desenvolvido por Broome (1999), isso significa que o escopo dos requisitos racionais é amplo, e não estreito.

O principal *insight* de Broome foi notar que a relação entre acreditar em algo e acreditar em sua consequência é outro tipo de relação, diferente daquela relação das "razões". Além disso, apesar de tentador, não podemos extrair de um requisito como "acreditar em *p* requer acreditar em *q*", ou seja, um requisito na forma "se *p*, então *q*", que temos razões para acreditar em *q*. Segundo Broome (1999, p. 406), a relação entre requisitos racionais é "estrita", enquanto a relação entre razões é "maleável". Na medida em que requisitos racionais se aplicam estritamente, eles são irrevogáveis e indiferentes a razões, pois possuem uma relação delimitada pelo escopo amplo do operador que governa aquela relação[11]. Por outro lado, "razões" são maleáveis pois são revogáveis frente a razões opostas, e possuem uma relação aberta com escopo estreito[12].

Algumas características são fundamentais para ser um requisito racional estrutural. Primeiramente, ele deve governar um conjunto de estados, não um estado individualmente. Isso é conhecido na literatura como uma distinção entre requisitos "globais" e "locais". Tradicionalmente, o trabalho em epistemologia está preocupado com a racionalidade, ou justificação, de crenças individualmente. Nas palavras de Christensen:

> [...] muito do trabalho em epistemologia tem focado em descrever o que uma crença individual possui que faz dela racional, ou, mais comumente, o que a faz justificada, ou uma instância de conhecimento (CHRISTENSEN, 2004, p. 08).

Por outro lado, como veremos no próximo tópico, boa parte dos trabalhos em epistemologia formal concentram-se, em grande medida, em requisitos globais sobre a racionalidade do conjunto de estados como um todo.

Como apontado anteriormente, uma característica importante de um requisito racional estrutural é que ele possui escopo amplo: ele se aplica ao condicional que governa a relação entre os conteúdos presentes nos respectivos estados, e não aos próprios estados individualmente. Em um caso de implicação material, ele governa o condicional como um todo. Como já foi dito, do fato de que p implica q, e eu acredito em p, não podemos derivar qualquer conclusão normativa sobre se eu devo ou não acreditar em q. Broome (1999, p. 404) chama a relação entre estados dentro de um requisito racional de "relação não-desvinculável" (do inglês, *non-detaching relations*). E essa impossibilidade de desvincular conclusões normativas de requisitos racionais é um dos grandes atrativos para a defesa de que requisitos racionais possuem escopo amplo.

Uma consequência de tomar requisitos racionais como possuindo escopo amplo é que as exigências são simétricas entre estados: se eu acredito que p, e que p acarreta q, mas não acredito em q, posso solucionar esse conflito de três modos distintos: acreditando em q, abandonando a crença em p, ou abandonando a crença que p acarreta q. Ou seja, dentro de um requisito racional, não existe hierarquia entre os conteúdos. Isso significa que requisitos racionais são indiferentes ao conteúdo

[11] Por exemplo, suponha que R seja o operador modal para "A racionalidade requer que" e $Bel(p)$ represente o estado de crença em uma determinada afirmação p. Um requisito racional de escopo estreito toma a relação de implicação material entre p e q da seguinte forma: $\vdash Bel(p) \rightarrow R\ Bel(q)$. Por outro lado, um requisito de escopo amplo possui a seguinte forma: $\vdash R(Bel(p) \rightarrow Bel(q))$. Para mais detalhes, ver Broome (1999).

[12] Uma excelente exposição desse debate pode ser encontrado em Brunero (2010).

de cada crença particular. Resumindo, daqui em diante, requisitos racionais serão entendidos da seguinte forma:

DEFINIÇÃO 3. Seja ϕ uma norma da racionalidade, ϕ é um requisito racional estrutural *sse* ϕ é *global* e de *escopo amplo*.

Para um defensor da tese estruturalista, toda norma da racionalidade ou é um requisito racional estrutural, ou pode ser reduzido a um conjunto de requisitos estruturais.

Um dos principais desafios para o defensor dos requisitos racionais estruturais é seu caráter simétrico: dado uma contradição na forma *p* e *não-p*, posso evitar a irracionalidade abandonando qualquer um dos estados. Segundo Niko Kolodny (2007), isso traz consequências extremamente contra-intuitivas: um agente racional pode cumprir o requisito de não-contradição e, ainda assim, fazer isso contrário a suas razões; ele pode solucionar o conflito rejeitando a crença que possui maior respaldo evidencial. Nas palavras do próprio autor,

> [...] alguém torna sua crença consistente, mas 'contrário a razões'. Deixa-se de acreditar em *p*, o que, à luz das evidências, é suficientemente provável ser verdadeira e exigida pelas razões, enquanto continua acreditando em *não-p*, o que, à luz das evidências, é suficientemente provável ser falsa e proibida pelas razões (KOLODNY, 2007, p. 237).

Como um defensor da tese substancialista da racionalidade, a proposta de Kolodny é explicar o requisito de não-contradição a partir de uma teoria do erro: se o agente acredita em *p* e *não-p*, ou ele acredita em *p* com razões suficientes, ou ele acredita em *não-p* com razões suficientes. Ou seja, por trás de um requisito racional, existe sempre uma relação entre razões que motiva sua aceitação. Para o autor, a simples coerência exigida pelos requisitos racionais não possui qualquer valor significativo.

Por enquanto, vou deixar esse desafio em aberto. No final deste capítulo voltaremos a essa discussão, onde ofereço uma breve resposta (ainda não apresentada na literatura relevante) para tal questão. Partindo da literatura sobre a relação entre a racionalidade de conjuntos de crenças e conjuntos de probabilidade subjetiva, pretendo mostrar como a formulação apropriada de um Princípio Ponte pode oferecer novos, e interessantes, avanços na compreensão do debate entre substancialistas e estruturalistas.

1.2 FORMALIZANDO A RACIONALIDADE ESTRUTURAL DOS ESTADOS DOXÁSTICOS

No tópico anterior, a partir do conflito entre racionalidade estrutural e substancial, vimos que uma noção interessante de racionalidade encontra-se formulada sob o domínio de requisitos racionais. Contudo, a literatura até então abordada, que concentra parte significativa desse debate, está primariamente focada em questões de natureza prática. Dentro dessa literatura relevante, até o momento presente, não encontramos projetos já em nível avançado sobre a abrangência e aplicações de requisitos racionais[13]. Todavia, não é nenhuma novidade dizer que o tema racionalidade sempre foi objeto de investigação dos teóricos do conhecimento.

Podemos dizer que epistemólogos tradicionais se preocupam, entre outras coisas, com o debate em torno da racionalidade da crença. Dentre as preocupações sobre esse tema, encontramos principalmente discussões sobre o modo como podemos (ou devemos) entender a coerência entre crenças, como acomodar novas evidências, e sobre o papel da crença racional na tomada de decisão. Para além da preocupação tradicional dos epistemólogos pelo tema da racionalidade, encontramos um extenso debate sobre a racionalidade da crença no que conhecemos hoje por Epistemologia Formal. Epistemólogos formais se preocupam, em grande medida, com os mesmos objetos de estudo dos epistemólogos tradicionais, só que fazem isso utilizando-se de modelos matemáticos para investigar os fenômenos relevantes[14].

É possível afirmar que grande parte da literatura em Epistemologia Formal sobre a racionalidade da crença concentra-se no desenvolvimento de requisitos racionais estruturais, mas não exatamente nestes termos. Requisitos racionais para crença e probabilidade subjetiva, como serão apresentados abaixo, são desenvolvidos como uma teoria dos padrões formais para um conjunto de estados que se pretenda racional. Isso não diz nada sobre os conteúdos desses estados, nem possui qualquer implicação sobre o conteúdo para o qual possuímos mais ou menos razões para crer. É viável afirmar que é nos trabalhos em Epistemologia Formal da crença que encontramos o mais extenso desenvolvimento sobre requisitos racionais estruturais. Os princípios apresentados a seguir são, fundamentalmente, requisitos racionais de coerência entre estados, tais como exigidos pela noção estrutural de racionalidade[15].

1.2.1 Requisitos Racionais para Crença

Vamos agora começar a tornas as coisas um pouco mais precisas. Tomemos o objeto da crença como sendo “proposições”. Podemos afirmar que crenças e proposições possuem uma relação bem próxima: chamamos de proposição o conteúdo informacional de uma crença, o elemento da crença portador de valor de verdade. Como quase tudo em filosofia, não há consenso na literatura sobre o que exatamente

[13] Em Broome (2013), o autor apresenta um trabalho bastante abrangente sobre os requisitos racionais, mas focado principalmente em requisitos da racionalidade prática. Como já dito, nosso foco será nos casos teóricos.

[14] Para o leitor interessado no papel da construção de modelos formais na representação de fenômenos epistêmicos, recomendo a leitura do Apêndice A, onde eu exploro um pouco mais essa relação entre epistemologia formal e como um epistemólogo pode se beneficiar do uso de ferramentas matemáticas no desenvolvimento de seu trabalho analítico.

[15] O modo como Broome (1999, 2013) formula os requisitos racionais possui uma estrutura distinta do que veremos abaixo. Isso ocorre fundamentalmente em virtude de Broome não ser um epistemólogo formal, e ter preocupações distintas das abordadas aqui.

são "proposições". Todavia, o modo mais tradicional de entender o termo é por meio da noção de "mundos possíveis": proposições são conjuntos de mundos possíveis (deixarei isso mais claro à frente). De modo semelhante, também não há consenso sobre como compreender "mundos possíveis". Ainda assim, podemos dizer que mundos possíveis são especificações completas de todas e somente aquelas características do mundo[16].

De um modo bem intuitivo, falamos de proposições para se referir ao *sentido* de sentenças descritivas, e é por meio dessas sentenças que expressamos nossas crenças. Por exemplo, as sentenças "a neve é branca" e "*snow is white*" são obviamente distintas. Contudo, elas expressam a mesma proposição, isto é, elas possuem o mesmo sentido, apenas estão em idiomas distintos. Do mesmo modo, a sentença "não é o caso que a neve não é branca" expressa a mesma proposição que as duas anteriores. De um modo um tanto quanto informal, todas essas sentenças dizem a mesma coisa: elas são verdadeiras no mesmo conjunto de mundos possíveis. Agora de um ponto de vista mais formal, dizemos que proposições, enquanto conjunto de mundos possíveis, possuem uma estrutura de conjuntos. E é precisamente a partir de uma teoria de conjuntos que apresentaremos o que será aqui chamado de uma *Lógica* da crença.

Começando pelo espaço amostral, vamos denotá-lo por W. Ele irá conter todos os mundos possíveis relevantes. Portanto, W é compreendido como o conjunto de mundos possíveis que contém todas as proposições. Assim, proposições (subconjuntos de W) são conjuntos cujos elementos são mundos possíveis, e serão denotados por letras maiúsculas A, B, C, ..., etc. Denotaremos uma coleção de n mundos possíveis por $\{w_1, w_2, \dots, w_n\}$, com w_i sendo qualquer um dos n mundos, sendo n um número natural. Estes são modos como o mundo, ou parte relevante dele, pode ser. Mais precisamente, proposições, mundos possíveis, e verdade ou falsidade das proposições, são definidas como a seguir:

DEFINIÇÃO 4. Seja W um conjunto (não-vazio) de mundos possíveis:

1. A é uma proposição *sse* A é um subconjunto de W;

2. Se A é uma proposição e w_i um mundo em W, então A é verdadeira em w_i *sse* w_i é elemento de A;

3. Se A é uma proposição e w_i um mundo em W, então A é falsa em w_i *sse* w_i é elemento de $\neg A$ (ou $W \setminus A$)[17].

[16] Para maiores detalhes sobre esse debate, ver, por exemplo, o tópico *The Objects of Belief and their Structure* em Genin e Huber (2020). Esse verbete também contém uma apresentação bastante semelhante dos detalhes formais que serão apresentados logo abaixo.

[17] Por simplicidade de notação, utilizarei diversas vezes o símbolo lógico de negação $\neg A$ ao invés de A^c ou $X \setminus A$ para representar o complemento do conjunto A.

Formalmente, proposições podem ser expressas por meio de um *corpo de conjuntos*, ou uma *álgebra*. Tal álgebra (também chamada de *álgebra sobre W*) é formada a partir do conjunto das partes de W, ou seja, $\mathscr{P}(W)$. Denotaremos por $\mathcal{A}$ a álgebra dos subconjuntos de W, de modo que, se as proposições A e B forem subconjuntos de W e pertencerem à álgebra $\mathcal{A}$, a união $A \cup B$ também pertence a $\mathcal{A}$, assim como a intersecção $A \cap B$, e os respectivos complementos, $\neg A$ e $\neg B$. Desse modo, dizemos que $\mathcal{A}$ é fechada sob operações de união, intersecção e complemento (ou negação)[18].

Enfim, vamos aos requisitos racionais para crença. Os princípios que serão apresentados determinam padrões de inferências relativamente plausíveis. Primeiramente, é comum aceitar que crenças formam nossa imagem de mundo. Na maioria dos casos ela visa captar o mundo de um modo preciso. Alguns teóricos dizem que a crença "visa a verdade"[19]. Voltaremos a esse ponto em breve. Por hora, podemos justificar tais requisitos pelo seu caráter preservador de verdade *a priori*: independente de qualquer informação que possuo sobre qual dos mundos possíveis é o mundo atual, tais requisitos preservam a verdade das crenças sob sua jurisdição. Por exemplo, para qualquer proposição A, minha crença na tautologia $A \vee \neg A$ será verdadeira, pois para qualquer mundo w_i em W, ou w_i pertence a A, ou w_i pertence a $\neg A$. Isso faz com que minha crença, antes de qualquer conhecimento sobre o mundo, atinja seu alvo: a verdade. O oposto ocorre com as contradições: minha crença na contradição $A \wedge \neg A$ será falsa, pois para qualquer mundo w_i em W, ou w_i pertence a A, ou w_i pertence ao complemento de A, e nunca aos dois ao mesmo tempo.

A lógica por trás dos princípios que serão apresentados a seguir é ponto de partida de grande parte da literatura formal sobre racionalidade da crença. As principais referências aqui se encontram nos trabalhos de Lógica Epistêmica e Teoria de Revisão de Crença, mais precisamente, no cânone dos trabalhos de Lógica Epistêmica de Jaakko Hintikka (1962), e no trabalho fundante da teoria AGM do trio Carlos E. Alchourrón, Peter Gärdenfors e David Makinson (1985)[20]. Daqui em diante, $Bel(\cdot)$

[18] Ao longo da Tese vou estender para proposições a terminologia lógica que é normalmente utilizada para fórmulas e sentenças, sempre que isso tornar a leitura mais agradável. Por exemplo, o conceito "verdade lógica", normalmente atribuído a sentenças e fórmulas, irá se referir proposição W, ou seja, um conjunto contendo todos os mundos possíveis, enquanto o conceito de "contradição lógica" irá se referir ao conjunto vazio $\emptyset$. O mesmo acontecerá com o conectivo lógico de negação, disjunção, conjunção, etc.

[19] Para uma excelente exposição sobre esse tema, ver Chan (2013).

[20] Algumas considerações mais técnicas são importantes aqui. Na Lógica Epistêmica, a linguagem objeto é ampliada por meio de operadores epistêmicos, de modo que "S acredita que p" é expressa na linguagem pela fórmula SBp ou do modo mais comum $Bel_S(p)$. Em Hintikka (1962), uma semântica para tais operadores é desenvolvida através de mundos possíveis. Aqui me aproximo de Gärdenfors (1988) e concentro meus recursos em sistemas de crença, ou conjuntos de crenças, uma vez que nosso objeto de estudo não será operadores epistêmicos em uma linguagem natural ou formal, mas sim a estrutura do sistema de crenças do agente no que diz respeito a sua racionalidade. Nesse sentido, em termos sintáticos, um conjunto de crenças pode ser entendido também como uma *Teoria*, tal como compreendida pelos lógicos. Isso torna nossa linguagem objeto mais simples, sem

denotará uma atitude (ou operação) de crença sobre os membros da álgebra $\mathcal{A}$, e [**B**] denotará o conjunto de crenças do agente em questão, ou seja, o conjunto de proposições acreditadas[21]. Para dizer que uma determinada proposição A é acreditada pelo agente, podemos escrever $A \in [\mathbf{B}]$ ou como $Bel(A)$. A segunda opção será a mais utilizada aqui inicialmente. Por fim, vamos assumir que [**B**] possui um número finito de elementos.

O primeiro requisito garante a racionalidade da crença em uma proposição tautológica W, onde W é o conjunto de todas as possibilidades consideradas naquele momento:

$$Bel(W) \tag{1}$$

Uma vez que W contém todo espaço amostral, a proposição W é verdadeira em todos os mundos possíveis. Se a crença no maior conjunto W é uma crença tautológica, podemos imaginar que a crença em uma proposição contraditória é a crença em um subconjunto vazio, ou seja, ela não é verdadeira em nenhum mundo possível. Mais precisamente, uma determinada proposição A e seu complemento não podem ser ambas verdadeiras no mesmo mundo: não é possível existir um mundo w_i que é elemento de A e também é elemento de $\neg A$. Assim, a seguinte norma exige que um agente racional não creia em contradições[22]:

$$\text{não é o caso que } Bel(\emptyset) \tag{2}$$

Tal como uma Lógica para crença, também temos um princípio de implicação. Lembremos que se A é um subconjunto de B, então B é consequência lógica de A. Em outras palavras, se $w_i \in A$, então $w_i \in B$, e, consequentemente $A \models B$. Para um conjunto de crenças, dizemos que é racional acreditar na consequência lógica desse conjunto. Formalmente, para quaisquer duas proposições A e B:

$$\text{Se } Bel(A) \text{ e } A \subseteq B, \text{ então } Bel(B) \tag{3}$$

Por fim, assumimos que crença é fechada sob intersecções finitas, por um princípio conhecido por "aglomeração", ou princípio "conjuntivo". Esse requisito é similar ao anterior, a diferença é que agora um não precisa ser subconjunto do outro, basta

a necessidade do uso de operadores epistêmicos mais sofisticados.

21 No segundo capítulo também iremos interpretar $Bel(A)$ como a atribuição de uma atitude específica '*Bel*' sobre uma determinada proposição A. Ou seja, *Bel* será uma das possíveis atitudes doxásticas disponíveis que determinado agente racional deve adotar em relação a proposição A. Mas isso é assunto para o próximo capítulo.

22 Nos trabalhos de Lógica Epistêmica, tal princípio é aceito somente quando o operador modal de acessibilidade é *serial*. Para a teoria AGM, tal postulado é válido para qualquer conjunto não-trivial (não-absurdo) de crença. Os outros postulados apresentados aqui são válidos para quaisquer sistemas modais para crença e também para revisão na teoria AGM. Para mais detalhes, ver Hintikka (1962) e Gärdenfors (1988, p. 24).

apenas que eles compartilhem um mundo relevante em comum. Para quaisquer proposições A e B:

$$\text{Se } Bel(A) \text{ e } Bel(B), \text{ então } Bel(A \cap B) \tag{4}$$

Resumidamente, esses quatro princípios determinarão a racionalidade ou não de determinado conjunto de estados[23]. Essa combinação de princípios é também conhecida como a exigência de *fechamento dedutivo* e *consistência lógica* para o conjunto de crenças. Isso significa, respectivamente, que é racional acreditar em tudo que o conjunto de crenças acarreta, e que não é racional acreditar em contradições. Em outras palavras, [**B**] deve ser grande o suficiente para conter todas suas consequências lógicas, mas não tão grande ao ponto de permitir contradições. Isso é basicamente o que é assumido em Lógica Epistêmica desde o trabalho de Hintikka (1962), e em grande parte dos trabalhos sobre revisão de crença, como veremos com mais detalhes no Capítulo 4.

Agora, partindo desse conjunto de suposições básicas para a racionalidade da crença, temos o seguinte resultado que nos será útil ao longo da tese:

Proposição 1.2.1. *Seja W um conjunto finito de mundos possíveis. A partir dos itens (1), (2), (3) e (4) apresentados, para cada conjunto não-vazio de mundos possíveis expresso em W, existe uma proposição, denotada por B_W, tal que:*

$$\textit{Para toda proposição } A \subseteq W, \ Bel(A) \textit{ é racional sse } B_W \subseteq A.$$

Isso significa que B_W representa a menor descrição completa de como o agente acredita que o mundo atual seja: a menor conjunção de tudo que o agente acredita racionalmente. Ao longo do texto, esse resultado aparecerá diversas vezes, e no Capítulo 3 ele receberá uma atenção especial, onde veremos mais detalhes sobre suas consequências teóricas.

Antes de encerrar esse tópico sobre a racionalidade da crença e passar para o tema da racionalidade da probabilidade subjetiva, é importante fazer uma consideração. Até agora, vimos requisitos racionais que se aplicam de modo *sincrônico*: eles governam o sistema de crença de uma maneira estática, de modo que nenhuma nova crença é adicionada ou excluída durante esse processo. Traçando uma analogia, esses requisitos racionais são como uma "foto" do conjunto de estados. Mas também podemos ver tal conjunto como um "filme": podemos formular requisitos racionais que regulam

[23] Aqui o leitor pode se questionar se tais regras para a racionalidade da crença são realmente requisitos racionais estruturais. Tradicionalmente, como vimos, requisitos racionais (com escopo amplo) são formulados condicionalmente: é racional que (se p então q). Nesse sentido, entendo que os princípios apresentados aqui poderiam ser entendidos como a Lógica, ou o modelo formal, dos requisitos racionais estruturais. Assim, na medida em que todos os requisitos estruturais estariam fundamentados por essa lógica, podemos dizer que as quatro regras, quando tomadas em conjunto, determinam todos os requisitos racionais estruturais para crença.

o sistema de crença de modo *diacrônico*, ou seja, que comendem racionalmente a dinâmica do "entra e sai" de proposições no sistema ao longo do tempo.

Uma maneira de compreender tal distinção entre requisitos racionais é a partir do que alguns autores "requisitos de estado" e "requisitos de processo". Nas palavras de Kolodny,

> Devemos, por consequência, distinguir entre 'requisitos de estado", que simplesmente excluem estados nos quais alguém possui atitudes conflitantes, e 'requisitos de processo', que dizem como, daqui em frente, alguém deve formar, manter, ou revisar suas atitudes para evitar, ou escapar, de tais estados conflitantes (KOLODNY, 2005, p. 517).

Como veremos no tópico seguinte, no caso dos requisitos racionais para probabilidade subjetiva, temos um princípio bastante conhecido que cumpre esse papel: o princípio de condicionalização. Contudo, no caso das crenças, isso é um pouco mais complexo.

Sem adentrar especificamente nos detalhes, existe uma grande controvérsia sobre a existência e o modo mais apropriado de entender requisitos diacrônicos para a racionalidade da crença. Especialmente no Capítulo 4, aprofundaremos esse debate a partir de dois modelos formais que disputam espaço no debate sobre revisão de crença, são eles a teoria AGM de revisão de crença, e a teoria do raciocínio não-monotônico. Contudo, além dessas teorias formais para a revisão de crença, alguns autores sustentam que existe uma norma diacrônica mais elementar governando esse processo dinâmico. Segundo Broome (2013, p. 185), tudo leva a crer que existe um requisito racional de persistência para crença, que poderia ser formulado do seguinte modo:

DEFINIÇÃO 5 (Persistência da Crença)**.** Para qualquer tempo t anterior a t', se o agente acredita na proposição A em t, e nenhuma evidência relevante sobre A é adquirida entre t e t', então é racional que o agente acredite em A em t'.

Para o autor, a justificativa para tal requisito é que, com o passar do tempo, nossas crenças adquiridas desempenham papel importante nas mais diversas tarefas, sejam elas teóricas ou práticas, e crenças excluídas arbitrariamente seriam irracionais de um ponto de vista dinâmico, uma vez que gerariam uma incoerência temporal. Por hora, podemos apenas assumir que esse princípio possui considerável respaldo intuitivo. Não vou entrar, por enquanto, em detalhes sobre esse requisito e sobre as controvérsias que ele gera. Esse tema retornará com a devida atenção no último capítulo da Tese.

1.2.2 Requisitos Racionais para Probabilidade Subjetiva

Vamos agora olhar mais precisamente para o outro lado da moeda: vamos determinar os requisitos racionais para a versão quantitativa da "crença" (entendida aqui como sinônimo do termo mais amplo "atitude doxástica"). É geralmente bem aceito na literatura em Epistemologia Formal que o lado quantitativo da crença, também chamado de "probabilidade subjetiva", corresponde àquele estado conhecido por meio da Teoria da Probabilidade Subjetiva, desenvolvida fundamentalmente por Frank P Ramsey (1926) e Leonard Savage (1972). Essa noção surge basicamente a partir da ideia de que existem algumas medidas de probabilidade que representam crenças (expectativas ou simplesmente graus de confiança) do agente em questão. Uma das motivações iniciais dessa visão foi medir incertezas do agente em uma situação de tomada de decisão. Foi principalmente essa noção de crença como probabilidade subjetiva que deu origem ao que hoje conhecemos como Epistemologia Bayesiana[24].

Atualmente, Epistemologia Bayesiana é entendida como uma área preocupada com questões relacionadas ao conhecimento que se utiliza de três elementos essenciais: (i) probabilidade subjetiva como o estado doxástico relevante; (ii) axiomas da teoria da probabilidade; (iii) e princípio de condicionalização. Falarei a seguir um pouco mais sobre cada um desses elementos. Começando pelo primeiro, podemos afirmar que a probabilidade subjetiva é o concreto do edifício bayesiano. Mas como podemos medir algo subjetivo como o grau de crença do agente? Teóricos que hoje são considerados os precursores do bayesianismo, como Ramsey e Savage, aceitam que probabilidades subjetivas são exteriorizadas por meio de nossas escolhas práticas, e, de um modo bem conhecido, por nossa disposição de aposta sobre os mais diversos conteúdos. Por exemplo, ao realizar uma aposta, se você notar que sua possibilidade de vitória for pequena, isto é, se sua confiança naquele resultado específico estiver em patamar baixo, você fará a aposta somente se o valor a ganhar for muito maior que o valor a perder.

Para um epistemólogo bayesiano, uma condição para um agente epistêmico ser considerado racional é que seu conjunto de probabilidades subjetivas esteja em acordo com a teoria da probabilidade. Sendo mais preciso, a teoria da probabilidade contemporânea padrão é o resultado do trabalho de axiomatização do russo Andrei Kolmogorov (1956). Resumidamente, a probabilidade de um evento é medida a partir de uma escala real que vai de 0 a 1. Mais precisamente, chamemos de *espaço de probabilidade* a tripla ordenada $\langle W, \mathcal{A}, P\rangle$, onde W é o espaço amostral, $\mathcal{A}$ a álgebra relevante, e P é uma função de distribuição de probabilidades sobre os elementos de W. Tradicionalmente, em teoria da probabilidade, subconjuntos de W são entendidos como "eventos". Aqui, W denotará o conjunto de todos os mundos possíveis, ou seja, a disjunção de todas proposições em $\mathcal{A}$. Seguindo a mesma linha, tomaremos a função de probabilidade P como sendo uma *função de distribuição* de probabilidade subjetiva sobre os elementos em W. Portanto, assim como $Bel(\cdot)$, a função $P(\cdot)$ mede o grau de

[24] A partir daqui, e ao longo de todo o restante do trabalho, vou assumir que o leitor possui alguma familiaridade com a literatura introdutória padrão sobre bayesianismo. Isso pode ser encontrado em manuais como Bradley (2015) e no excelente trabalho de Michael Titelbaum (2022a,b), volume 1 e 2.

crença do agente em relação a uma determinada proposição.

Tal como desenvolvida por Kolmogorov (1956), a teoria da probabilidade contemporânea consiste em três axiomas e uma definição de probabilidade condicional. O primeiro deles diz que, para qualquer evento, ou conjunto de eventos, a menor probabilidade possível é 0. Seja A um subconjunto de W:

$$P(A) \geq 0. \tag{5}$$

Este é conhecido como axioma da *Não-negatividade*.

De modo não surpreendente, o segundo axioma garante que a maior probabilidade que temos é igual a 1, tal que W tenha probabilidade máxima. Este é o axioma da *Normalização*:

$$P(W) = 1. \tag{6}$$

O terceiro axioma diz como devemos combinar a probabilidade de diferentes eventos. Se os eventos A e B possuem uma intersecção vazia, ou seja, não pode ser o caso que os dois ocorram simultaneamente, a probabilidade da união de A e B (quando um ou outro ocorre) será igual a soma da probabilidade individual de A e B. Este é o axioma da *Aditividade*:

$$P(A \cup B) = P(A) + P(B) \text{ } se \text{ } P(A \cap B) = \emptyset. \tag{7}$$

Nesse caso, também podemos chamá-lo de axioma da aditividade *finita*[25].

Além dos três axiomas apresentados, temos também uma definição de *Probabilidade Condicional*. Se queremos saber qual a probabilidade do evento A acontecer dado a ocorrência do evento B (denotado por $A \mid B$), seguimos a seguinte definição:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{(sendo } P(B) \neq 0\text{)} \tag{8}$$

Essa definição de probabilidade condicional determina que, ao assumir uma proposição como verdadeira, é como se todo o nosso universo W se restringisse apenas aos mundos compatíveis com àquela proposição assumida. Assim, dado B, a probabilidade de A deve se adaptar a esse novo cenário: um mundo onde B é tomado como certo.

Os axiomas de não-negatividade, normalização, aditividade, mais a definição de probabilidade condicional, constituem o fundamento da teoria da probabilidade contemporânea. Disso derivamos várias outras regras (teoremas) probabilísticas, e muitas

[25] Caso W seja um conjunto com um número infinito de mundos, precisamos de uma álgebra $\mathcal{A}$ que seja fechada sob intersecções *contáveis*. Nesse caso, nossa álgebra será uma $\sigma -$ *álgebra*, e chamamos essa versão de axioma da Aditividade *Contável*, ou $\sigma -$ *Aditividade*:

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), \text{ se } A_i \cap A_j = \emptyset \text{ para } i \neq j.$$

destas serão apresentadas ao longo deste trabalho. Portanto, ao equiparar a função P de probabilidade de Kolmogorov como sendo uma função de probabilidade subjetiva, podemos dizer que tais axiomas, em conjunto com suas consequências, garantem a existência de um conjunto de requisitos racionais quantitativos: eles garantem a coerência do conjunto doxástico de probabilidades subjetivas. Portanto, apresentamos até aqui 2/3 do que é o bayesianismo: (i) probabilidade subjetiva como o estado doxástico relevante; e (ii) axiomas da teoria da probabilidade como requisitos racionais para tais estados. Falta ainda apresentar o terceiro pilar do bayesianismo: o princípio de condicionalização.

Até o momento, falamos apenas de requisitos probabilísticos sincrônicos, ou seja, que não medem variações temporais. Contudo, diferentemente do caso das crenças qualitativas onde há pouco consenso sobre padrões dinâmicos de revisão de crença, com a probabilidade subjetiva temos certo consenso em torno de qual princípio de atualização é o mais apropriado. Há, na epistemologia bayesiana, um conhecido princípio que diz como atualizar a probabilidade subjetiva de uma proposição quando nova evidência é adquirida no instante seguinte. Esse é o famoso *Princípio de Condicionalização*:

DEFINIÇÃO 6 (Princípio de Condicionalização (Estrita))**.** Para qualquer tempo t anterior a t', se o agente recebe a evidência, em um instante entre t e t', de que é o caso que B, e $P(B) \neq 0$, então:

$$P_{t'}(A) = P_t(A \mid B).$$

Este é um princípio bayesiano diacrônico, ou seja, ele estabelece uma norma dinâmica para atualização do conjunto de probabilidades subjetivas do agente[26]. É importante mencionar que a fundamentação matemática da teoria da probabilidade de Kolmogorov é estática, ela não diz nada sobre atualização temporal. O princípio de condicionalização é fundamentalmente filosófico/epistêmico, o que o torna ainda mais atraente para nós.

Uma das características do princípio de condicionalização é que ele é definido apenas quando a evidência adquirida é tomada como certa, ou seja, como possuindo probabilidade 1. Isso ocorre, como vimos, pela natureza da definição de probabilidade condicional. Como consequência, a condicionalização estrita cria e retém certezas: a partir do momento que uma nova proposição com probabilidade 1 é adicionada ao conjunto de probabilidades subjetivas, ela nunca mais perde seu *status* de certeza. Mas alguém poderia questionar se isso não é uma exigência muito forte, e se um princípio diacrônico não deveria dar conta dos casos onde a nova informação adquirida não possui probabilidade máxima. A resposta mais aceita a esse desafio foi oferecida por

[26] Compare-o com a noção de “requisitos de processo” e com a definição de Persistência para Crença, apresentada anteriormente na Definição 5.

Richard Jeffrey (1990). Sua proposta foi tão bem recebida que o novo princípio de condicionalização, agora definido para qualquer valor de probabilidade para a nova evidência adquirida, recebeu seu nome:

DEFINIÇÃO 7 (Condicionalização de Jeffrey)**.** Para qualquer tempo t anterior a t', seja $B_1, B_2, \ldots, B_n$ uma partição de W cujos elementos possuam probabilidade positiva, e $\sum_i B_i = 1$. Seja A uma proposição qualquer: Se, em um instante entre t e t', o agente recebe nova evidência sobre B_i, onde $P_{t'}(B) \in (0, 1]$, então

$$P_{t'}(A) = \sum_{i=1}^{n} P_t(A \mid B_i) \cdot P_{t'}(B_i).$$

Quando $P(B) = 1$, a condicionalização de Jeffrey se reduz à condicionalização estrita. Nos casos mais simples, onde a partição é dividida apenas entre a proposição e seu complemento, ou seja, uma partição de apenas duas células, B e $\neg B$, a condicionalização de Jeffrey se resume a seguinte fórmula:

$$P_{t'}(A) = P_t(A \mid B) \cdot P_{t'}(B) + P_t(A \mid \neg B) \cdot P_{t'}(\neg B).$$

Concluída a breve apresentação do princípio de condicionalização, podemos dizer que os pilares do bayesianismo estão estabelecidos: a noção de probabilidade subjetiva, os três axiomas, mais a definição de probabilidade condicional, e o princípio de condicionalização (estrito ou de Jeffrey). Essa é a base da racionalidade bayesiana para formação e atualização de probabilidades subjetivas[27]. Nesse sentido, um epistemólogo bayesiano considera tais regras como critérios de correção: se um agente possui um conjunto de probabilidades subjetivas em desacordo com tais normas, dizemos que esse agente é epistemicamente inconsistente. Mais especificamente, dizemos que o agente que descumpre tais regras está em desacordo com os requisitos racionais para probabilidade subjetiva.

Vários teóricos tentaram, e ainda tentam, oferecer uma justificação para a racionalidade dos axiomas probabilísticos. Dentro da literatura bayesiana, a mais famosa das explicações está no conhecido argumento *Dutch Book*. Resumidamente, o argumento *Dutch Book*[28], em sua versão sincrônica, implica que um agente com graus de crença que desobedeçam aos axiomas da teoria da probabilidade pode ser pego em um conjunto de apostas (ou “livro” de apostas - vem daí o termo *book*) na qual ele certamente perderá dinheiro.

Esse ambiente de apostas sempre esteve muito conectado com a literatura de probabilidade, e por consequência, com a epistemologia bayesiana. Dentro de uma

[27] Os princípios que foram apresentados até então são relativamente consensuais dentro do bayesianismo, e podem ser facilmente encontrados em formulações parecidas em manuais de epistemologia bayesiana, como mencionados na nota 24.

[28] Uma das primeiras formulações desse argumento foi apresentado por Frank P Ramsey (1926).

visão bayesiana padrão, dizemos que a probabilidade subjetiva de um agente em uma proposição *A* pode ser medida por meio de sua *razão de aposta* (do inglês, *betting ratio*), isto é, o preço mais alto que o agente está disposto a pagar por um bilhete de aposta que paga R$1 se *A* for verdadeira, e R$0 se *A* for falsa. Esse seria um método que os primeiros proponentes do bayesianismo desenvolveram para tentar explicar como a racionalidade de algo tão subjetivo como "crenças" (nesse caso, probabilidades subjetivas) pode ser avaliada de um ponto de vista de terceira pessoa.

Para um aprofundamento maior desse debate sobre a justificação dos axiomas probabilísticos por meio de um *Dutch Book* seriam necessárias várias outras considerações e especificações que fogem ao escopo do presente trabalho. Por exemplo, pessoas possuem diferentes inclinações ao risco de perda monetária, o que dificultaria o trabalho de encontrar um valor exato para a probabilidade subjetiva de cada agente em questão. Mas tudo isso já foi bastante explorado pela literatura, e não terá grande relevância agora[29]. No Capítulo 4 retomaremos essa discussão com um pouco mais de detalhes.

O que irá nos interessar especificamente, são dois problemas que aparecem com esse tipo de justificação para o bayesianismo. O primeiro deles diz respeito à dificuldade em se oferecer uma versão diacrônica do argumento *Dutch Book* para o princípio de condicionalização[30]. O segundo, e mais relevante aos propósitos do capítulo atual, é o fato de que esse tipo de argumento repousa sobre uma condição pragmática de avaliação: a perda de bem estar do agente, refletido por meio do valor monetário da aposta. Ou seja, a justificação tradicional do bayesianismo é pragmática, e isso é amplamente reconhecido pela literatura. Contudo, esse debate ganhou novos desdobramentos a partir da publicação do artigo *A nonpragmatic vindication of probabilism* de James Joyce (1998). Esse texto é reconhecido por conter uma das primeiras tentativas de justificar os axiomas probabilísticos de um ponto de vista puramente epistêmico. É sobre isso que falaremos no tópico seguinte.

1.2.3 Requisitos racionais epistêmicos devem visar verdade/acurácia

Quando o assunto é racionalidade epistêmica, queremos que nossos estados doxásticos sejam mais adequados ao mundo, em outras palavras, queremos que o conteúdo ali presente seja verdadeiro. Partindo de uma suposição geral sobre a racionalidade de probabilidades subjetivas, Joyce (1998) então desenvolveu um meio de avaliação epistêmica para tais estados na forma de uma "medida de acurácia". Resumidamente, acurácia é tomada como uma medida de distância da verdade, e uma medida de acurácia é uma forma de avaliar como probabilidades subjetivas aproximam-se da

[29] Ver por exemplo, o capítulo 09 em Titelbaum (2022b) e o capítulo 03 em Bradley (2015) para uma visão panorâmica sobre o extenso debate em torno da justificação dos axiomas probabilísticos por meio do argumento *Dutch Book*.

[30] Ver, por exemplo, Christensen (1991).

verdade. Na filosofia, a posição dominante toma os conceitos de verdade e falsidade como categóricos, ou seja, não possuem graduações (não existe “meia verdade”!). Vamos lembrar que, pela definição de verdade de uma proposição, A é verdadeira em um mundo w_i quando $w_i \in A$. Por consequência, medimos a acurácia de uma probabilidade subjetiva pela distância entre o seu valor, pertencente ao intervalo $[0, 1]$, e o valor de verdade da proposição em questão. Quando A for o caso, quanto mais a probabilidade subjetiva se aproximar do valor máximo, mais acurada será, e, quando A for falsa, quanto mais a probabilidade subjetiva se aproximar do valor mínimo, mais acurado será o estado.

Vale ressaltar que a motivação inicial de Joyce era mostrar que um agente que possua probabilidades subjetivas que obedeçam aos axiomas da probabilidade serão sempre mais acurados do que os agentes que descumprem os axiomas, e isso acontece de uma maneira *a priori*, independente de qualquer informação que o agente possua sobre o mundo. Desse modo, Joyce ofereceu um argumento puramente epistêmico para requisitos racionais probabilísticos. No próximo capítulo, precisamente na Seção 2.3, retomarei o argumento da acurácia de Joyce, e será apresentado mais detalhadamente sua construção.

Podemos dizer que o argumento de Joyce mostra que os axiomas da teoria da probabilidade funcionam como um guia à verdade para estados doxásticos quantitativos. Como vimos anteriormente na Seção 1.2.1, o mesmo ocorre para requisitos racionais para crença: acreditar em uma proposição tautológica garante que o agente atinja a verdade (por mais trivial que isso seja) independentemente de como o mundo atual seja, e no caso das contradições, não existe um mundo possível onde A e $\neg A$ sejam verdadeiras simultaneamente, o que faz com que $A \cap \neg A = \emptyset$, ou seja, não existe um mundo possível que satisfaz tal condição, tornando assim tal proposição sempre falsa.

Embora grande parte da epistemologia tradicional sempre tenha aceito que a racionalidade exija da crença algo como a verdade, o modo como isso ocorre é bastante controverso. Na Seção 1.2.1, mencionamos que muitos autores sustentam que a verdade é o “alvo da crença”. Há inúmeras maneiras de entender essa afirmação. Por enquanto, vamos nos focar no caso da crença (*simpliciter*), que é o estado doxástico predominante na epistemologia tradicional, onde a discussão a seguir está situada. Depois veremos como isso se aplica às probabilidades subjetivas.

Quando questionados sobre a razão para os conceitos de crença e verdade possuírem tal vínculo, filósofos da mente, da linguagem, e do conhecimento, sustentam que o conceito de crença possui certo vínculo com a verdade, em grande parte, pelo fato de que crença é um estado mental que se diferencia de outros estados, como desejos e intenções, em razão de sua direção de ajuste: a crença possui uma direção de ajuste mente-mundo, enquanto desejos e intenções (também chamados de *pró-*

atitudes) possuem uma direção de ajuste mundo-mente[31]. No primeiro caso, temos um estado mental com a principal característica de captar, mapear, e descrever o mundo tal como ele é; já no segundo caso, temos estados mentais cuja principal característica é mover o agente em direção a alguma realização no mundo, fazendo com que o mundo se adapte a seus estados mentais, isto é, satisfazendo seus desejos e intenções.

Isso faz com que alguns autores defendam a existência de uma norma que conecte crença e verdade de alguma forma[32]. Allan Gibbard inicia um de seus artigos afirmando que "para crença, correção é verdade. Crença correta é crença verdadeira. Minha crença de que a neve é branca é correta somente no caso em que a crença é verdadeira, somente no caso em que a neve for branca" (GIBBARD, 2005, p. 338). Na tentativa de captar tais alegações, poderíamos esboçar um princípio da seguinte forma:

DEFINIÇÃO 8 (Norma de Correção para Crença (*NCC*))**.** Uma crença é correta *sse* a proposição acreditada é verdadeira.

O problema com essa formulação está no modo como devemos compreender o que significa para uma crença ser "correta". Se correta significa "ser verdadeira", esse princípio é claramente trivial. Se por "correto" queremos dizer "racional", isso entra em conflito um princípio da racionalidade apresentado anteriormente (na Seção 1.1), que diz que a racionalidade sobrevém ao mental. Entendida desse modo, *NCC* faria com que toda crença verdadeira fosse racional, e toda crença falsa irracional, o que parece algo bastante estranho. Na epistemologia, "verdade" é claramente um tipo de propriedade não-mental, o que torna essa norma de correção para crença inconsistente com o que estamos tomando como sendo a racionalidade da crença.

Não é difícil prever um modo de reformular *NCC* para torná-lo apropriado. Uma vez que atingir ou não a verdade é algo que, na maioria dos casos, escapa ao nosso controle, podemos dizer que a racionalidade da crença não é diretamente dependente da verdade, mas sim, de buscar os melhores meios para atingi-la, em outras palavras, visar a verdade:

DEFINIÇÃO 9 (Norma de Racionalidade para Crença (*NRC*))**.** Uma crença é racional *sse* ela "visa a verdade".

Agora é necessário explicar o que exatamente é "visar a verdade". Uma resposta mais simples diria que visar a verdade é ter a verdade como alvo. Mas crenças não são flechas que podem ser apontadas para algo. Ter a verdade como alvo é uma metáfora.

[31] Um clássico exemplo dessa distinção aparece com o filósofo da mente e linguagem, John Searle (1983).

[32] Ver, por exemplo, Engel (2013) sobre as possíveis formulações (semânticas, metafísicas, epistêmicas) de uma norma que conecte crença e verdade. Em Chan (2013) encontramos uma coleção de artigos dedicados a essa discussão.

Segundo Wedgwood (2002), ter a verdade como alvo da crença é tomar os melhores meios disponíveis para alcançar o fim desejado. Mas o que seria esse meio? Nas palavras do autor,

> Suponha que você esteja agora tentando escolher qual regra seguir para atingir esse objetivo. Qual é o modo racional de escolher qual regra seguir? Uma vez que o mais longe que você pode ir desse objetivo é acabar acreditando em algo falso, você deve, presumivelmente, visar não seguir qualquer regra que, na circunstância, pode facilmente te conduzir, no fim, a acreditar em algo falso. (WEDGWOOD, 2002b, p. 276)

Suponha que você esteja tentando escolher (se isso for psicologicamente possível) quais regras seguir para atingir o objetivo de ter crenças verdadeiras. Qual é a forma adequada de escolher quais regras seguir? Você deve evitar seguir qualquer regra que, em última instância, possa te conduzir facilmente a aquisição de crenças falsas. No caso da probabilidade subjetiva, a escolha racional é por regras que conduzam a um conjunto de estados mais acurados, e isso é basicamente o que faz Joyce com o argumento em defesa de uma justificação epistêmica para o bayesianismo. Assim, podemos dizer que requisitos racionais estruturais garantem, *a priori*, que um conjunto de estados doxásticos não se torne inconsistente, e por consequência, falhe em atingir a verdade.

Contudo, não é difícil notar que o tipo de exigência em *NRC* não possui a forma de um requisito racional: ele é local; de escopo estreito; assimétrico; e não é indiferente ao conteúdo. Resumidamente, *NRC* é local pois requer algo de cada proposição individualmente: que ela vise a verdade. O que não é o caso para requisitos racionais estruturais[33]. Se *NRC* não é um requisito racional, que tipo de coisa ele é? Uma opção seria dizer que *NRC* é mais uma dentre tantas normas da racionalidade, juntamente com os requisitos racionais estruturais. Mas isso parece algo difícil de se aceitar para um defensor da tese da racionalidade estrutural. Assumir a existência de requisitos racionais de escopo estreito, locais, e assimétricos, parece ir em direção oposta a visão de que racionalidade é sobre o equilíbrio entre estados (como viemos trabalhando desde o início)[34]. Portanto, uma concepção estrutural, tal como foi assumida aqui, parece

[33] Alguém poderia questionar que o requisito para crença $Bel(W)$ apresentado anteriormente exige algo de uma proposição individualmente: que tautologias sejam acreditadas. Contudo, é importante mencionar que cada uma das regras que fundamentam o que estamos chamando de Lógica da crença, quando tomadas individualmente, não possuem qualquer relevância. Os itens (1), (2), (3) e (4) da Seção 1.2.1 justificam-se como requisitos racionais na medida em que fazem parte de um conjunto de regras de inferência que garantem o equilíbrio do sistema como um todo. Assim, o requisito $Bel(W)$ parece exigir algo individualmente de uma crença (ou seja, que tautologias sejam acreditadas) na medida em que os outros requisitos, como o de consequência lógica sob operações de subconjunto, aplicam-se simultaneamente a todo sistema de crenças. O mesmo ocorre para os requisitos racionais para probabilidade subjetiva, na forma dos axiomas probabilísticos.

[34] Easwaran e Fitelson (2015) sustentam que, para além dos requisitos racionais estruturais, existem outras normas da racionalidade que possuem escopo estreito. Uma delas seria semelhante a *NRC*. Mas os autores sustentam também que existe uma outra norma para a racionalidade da crença que toma a evidência, e não a verdade, como seu principal alvo. E isso estaria refletido no debate entre

excluir a existência de um pluralismo sobre requisitos racionais.

Já outros teóricos, como Wedgwood (2002b), sustentam que *NRC* seria um tipo de norma epistêmica fundamental, de modo que todos os requisitos racionais seriam explicados a partir desse princípio fundamental que conecta crença e verdade. Nesse sentido, todos os requisitos racionais seriam "derivados", em algum sentido, de uma norma epistêmica fundamental que conecta crença e verdade, como faz *NRC*. Não vou adentrar aqui no debate sobre o que seria esse tipo de norma, e qual a sua natureza. Alguns autores sustentam que a existência de uma norma epistêmica fundamental, irredutível, nos comprometeria com algum tipo de norma *sui generis* sobre a racionalidade e natureza da crença[35]. Meu objetivo será mostrar que podemos preservar uma conexão importante entre crença e verdade sem expandir nossa ontologia de normas epistêmicas (o que, diga-se de passagem, considero digno de algum mérito).

Uma possibilidade, ainda não explorada pela literatura atual sobre o tema, que considero apropriada é simplesmente assumir *NRC* como parte da *definição* de racionalidade epistêmica. É prática comum na literatura relevante distinguir racionalidade epistêmica e prática. Além de diferenciarem-se por seus estados mentais resultantes de seus respectivos raciocínios (um raciocínio epistêmico conclui em uma crença, enquanto um raciocínio prático conclui com uma intenção[36]), e da direção de ajuste desses estados, racionalidade prática e teórica são fundamentalmente separados por seus fins. De um modo um tanto quanto sucinto, é possível definir racionalidade epistêmica da seguinte maneira:

DEFINIÇÃO 10 (Racionalidade Epistêmica)**.** Um estado doxástico é epistemicamente racional *sse* satisfaz requisitos racionais que visam a verdade (ou acurácia).

Tal definição pode ser relevante na medida em que conecta crença e verdade de um modo totalmente instrumental, sem apelar para qualquer tipo de norma exclusivamente epistêmica.

Por outro lado, embora não seja o alvo do presente trabalho, o conceito de racionalidade prática é geralmente conectado a "valores práticos", como utilidade, prudência, bem-estar, etc. Sendo assim, é possível definir racionalidade prática do seguinte modo:

DEFINIÇÃO 11 (Racionalidade Prática)**.** Uma intenção é praticamente racional *sse* satisfaz requisitos racionais que visam valores práticos (ex., utilidade, prudência, bem-estar, etc.)

concepções *aléticas* e *evidenciais*, presentes no clássico debate entre William James e William K. Clifford. Para Easwaran e Fitelson, uma posição substancialista como a de Kolodny (2005) repousa-se exatamente sobre uma norma de evidência, e não em uma norma de verdade.

[35] Papineau (2013) argumenta que a relação entre crença e verdade, se normativa, é de um modo puramente instrumental. Não existe qualquer norma irredutível, de natureza puramente epistêmica, para crença que exija que ela vise a verdade.

[36] Ver, por exemplo, Harman (1986), Bratman (1987) e Broome (2013).

É importante notar que, no caso da racionalidade prática, avaliamos sua racionalidade a partir de um estado, que é a intenção, e não de uma ação. Isso ocorre, fundamentalmente, em virtude da tese da sobreveniência da racionalidade sobre o mental, e uma ação é considerada uma atitude física, não mental[37]. E no caso da racionalidade epistêmica, note que me refiro ao conceito amplo de estado doxástico, que abrange tanto crenças quanto probabilidades subjetivas.

Uma noção clara de racionalidade epistêmica é importante para distinguir aqueles casos onde parece ser racional acreditar em algo, porém, por razões práticas. Um clássico exemplo na literatura é a *Aposta de Pascal*, onde encontramos um argumento em defesa da racionalidade da crença em Deus, mas a partir de um valor prático de prudência[38]. Uma resposta comum ao argumento da Aposta de Pascal é que ele não oferece nenhuma razão a favor da *verdade* da existência de Deus[39]. Portanto, sem uma distinção entre racionalidade epistêmica e prática, fica difícil desenvolver uma teoria da racionalidade que dê conta das intuições extraídas com esse tipo de caso.

Sabemos que crenças são geralmente bastante úteis, mas não é esse aspecto que nos interessa aqui. Nas palavras de Gibbard,

> Utilidade pura e simples para tal propósito, todavia, não é o único modo pelo qual a verdade da crença pode importar. Buscamos a verdade, algumas vezes, puramente para conhecê-la. [...] De fato, é tal busca desinteressada pela verdade, talvez, que subjaz à racionalidade de um tipo especial da crença, uma racionalidade não pragmática, mas puramente epistêmica. [...] Aceito que existe tal coisa como racionalidade puramente epistêmica, e que isso pode, algumas vezes, contrastar com a desejabilidade pragmática da crença, da qual é racional querer acreditar. (GIBBARD, 2007, p. 143)

É claro que isso não esgota o debate em torno da relação entre crença e verdade. Não é meu propósito aqui sustentar uma tese robusta a favor ou contra determinada teoria da racionalidade, o que exigiria um trabalho à parte. Meu objetivo nesta seção foi apresentar alguns dos desdobramentos teóricos de uma concepção estrutural da racionalidade aplicada à literatura em epistemologia formal sobre a racionalidade da crença e probabilidade subjetiva. Vimos que requisitos racionais estruturais podem desempenhar papéis relevantes dentro de uma visão importante de racionalidade, e que podemos explicar a relação entre os conceitos de “verdade” e “racionalidade” de um modo totalmente instrumental, a partir de uma distinção entre racionalidade prática e teórica. Voltaremos a esse assunto no próximo capítulo, onde o argumento da

[37] Um dos pioneiros dessa vertente de discussão sobre a natureza da racionalidade prática e epistêmica a partir dos respectivos estados mentais foi Gilbert Harman (1986). Esta obra é referência para boa parte do que estamos discutindo no presente tópico, e diria que uma das principais influências para o presente trabalho como um todo.

[38] Para mais detalhes sobre esse argumento, ver Hájek (2018).

[39] No primeiro capítulo de sua obra, Christensen (2004) apresenta brevemente essa distinção entre racionalidade epistêmica e prática e mostra como o caso da Aposta de Pascal é elucidativo para essa questão.

acurácia em defesa dos axiomas da teoria da probabilidade como exigências da racionalidade desempenhará papel central, e também retomaremos esse tópico no último capítulo, onde abordaremos a dificuldade de ser formular argumentos semelhantes para racionalidade epistêmica diacrônica.

1.3 QUAL A RELAÇÃO ENTRE CRENÇA E PROBABILIDADE SUBJETIVA?

Tudo o que vimos até então no presente capítulo é material acessório ao nosso principal desafio. Como já foi dito, o objetivo primário da presente Tese será apresentar e investigar detalhadamente algumas das possíveis respostas à seguinte pergunta: como devem, de um ponto de vista da racionalidade estrutural, se relacionar os conceitos de crença e probabilidade subjetiva? Veremos aqui que essa resposta não é nada trivial. No restante deste capítulo pretendo apresentar de modo mais preciso nosso problema central e mostrar quais são as opções de caminhos que podem ser tomados para tratar o desafio em questão.

1.3.1 Eliminativismo

Ao pensar sobre uma resposta adequada à pergunta do parágrafo acima, a primeira opção que pode vir à mente de alguns consiste em, simplesmente, rejeitar uma das visões como simplesmente equivocada, resultado de um mal uso da linguagem, ou fruto de teorias erradas. Parte dessas teses repousam sob um pressuposto psicológico sobre a natureza desses estados. Para eles, não existe um estado mental, em agentes humanos, que corresponde ao que é chamado aqui de crença, ou à assim chamada probabilidade subjetiva. Outra parte repousa sobre um pressuposto "normativo" (mais precisamente, não-descritivo) sobre a racionalidade desses estados. Para eles, mesmo que exista empiricamente uma prática comum partilhando de algum desses estados, um dos dois estados não possui qualquer padrão de avaliação que possa ser classificado como racional ou irracional, de modo que um dos conjuntos de requisitos racionais apresentados na Seção 1.2.1 e Seção 1.2.2 é simplesmente fruto de teorias equivocadas sobre a racionalidade humana.

É claro que, na prática, as divisões entre quais filósofos defendem o quê, não se encaixam sempre perfeitamente no enquadramento do parágrafo anterior. Além disso, a visão eliminativista é minoritária dentro da literatura sobre o tema. Richard Jeffrey (1970) é o maior representante de um dos lados desse conflito. Ele é considerado um tipo de bayesiano radical ao acreditar que a única maneira racional de representar estados doxásticos é por meio de probabilidades subjetivas. Para o autor, na medida em que o formalismo bayesiano desenvolveu meios objetivos de mensurar, por meio de um comportamento de aposta, isso que é tradicionalmente chamado pelos epistemólogos de crença, podemos abandonar a noção clássica de crença e colocar em seu lugar a concepção bayesiana de probabilidade subjetiva. Em um famoso trecho, Jeffrey afirma:

> [...] nossa noção ordinária de crença está presente somente nos vestígios da noção de graus de crença. Estou inclinado a pensar que Ramsey retirou todo o tutano da noção ordinária e o utilizou para alimentar uma visão mais adequada (JEFFREY, 1970, p. 172).

Outro autor que acredita que Ramsey extraiu toda a essência do conceito tradicional de crença em benefício de uma visão mais adequada, expressa por meio de probabilidades subjetivas, é David Christensen (2004). Para o autor, na medida em que os requisitos racionais para crença (*simpliciter*) implicam que o sistema de crenças do agente seja fechado dedutivamente, famosos contraexemplos a essa exigência (como veremos logo à frente) nos conduzem à rejeição do fechamento dedutivo para crenças, e por consequência, à rejeição dos requisitos racionais apresentados anteriormente na Seção 1.2.1[40]. Para o autor, o caráter intuitivo, e a aceitação amplamente compartilhada, do fechamento dedutivo como requisito racional para crença pode ser explicado simplesmente a partir dos requisitos probabilísticos, sendo este último, o único conjunto de requisitos racionais legítimos para estados epistêmicos.

Por outro lado, Gilbert Harman (1986) é conhecido como um dos maiores representantes do outro lado da moeda. Ele argumenta que probabilidades subjetivas são simplesmente epifenômenos de nossa fala sobre força de adesão a uma crença que tornam-se explícitos em uma situação dinâmica de revisão e abandono de crença (HARMAN, 1986, p. 22). Para o autor, probabilidades subjetivas não desempenham nenhum papel significativo em nosso estudo da racionalidade e raciocínio. Contudo, Harman não é apenas um eliminativista sobre probabilidades subjetivas, ele é um eliminativista sobre qualquer requisito racional sincrônico para crença. Para ele, uma teoria da racionalidade diz respeito à mudança de visão, que tradicionalmente chamamos de raciocínio. E nesse processo, a lógica clássica, tal como conhecemos, não pode desempenhar um papel central. Uma teoria da racionalidade deve levar em conta o aspecto *não-monotônico* do raciocínio humano, o que é incompatível com a monotonicidade da lógica clássica (falaremos mais sobre isso no Capítulo 4).

Por fim, não podemos deixar de mencionar outro reconhecido teórico em defesa da eliminação das probabilidades subjetivas como estados legítimos. Segundo Richard Holton (2014, p. 33), existe uma proximidade significativa entre estados de crença e intenção: ambos são qualitativos (binários), estáveis, e permitem a coordenação na ação. Para o autor,

> A abordagem bayesiana não é uma idealização de algo que realmente fazemos. Ao invés disso, é algo bastante estranho à nós. Assim como nosso

[40] No final do segundo capítulo, veremos que existe um modo de tentar justificar a rejeição do fechamento dedutivo sem abandonar todos os requisitos racionais para crença. Isso é desenvolvido fundamentalmente por Easwaran e Fitelson (2015).

> principal estado deliberativo nativo é simplesmente o de intenção, do mesmo modo nosso principal estado epistêmico nativo é simplesmente o de crença cheia (HOLTON, 2014, p. 15).

Assim como Harman, Holton também acredita que os casos de uso de probabilidades subjetivas oferecidos como argumento pelos bayesianos podem ser simplesmente interpretados como uma fala sobre o "conteúdo probabilísticos" das crenças, e não como um tipo distinto de atitude quantitativa. Assim, requisitos racionais probabilísticos não são apenas desnecessários, mas são confusões conceituais sobre o papel dos estados doxásticos na vida de agentes reais.

Dentro desses exemplos, existem divergências quanto à natureza psicológica desses estados, e sobre a existência ou não de um deles. Mas todos eles convergem sobre a existência de requisitos racionais de apenas um dos tipos relevantes de estados doxásticos. Essa não vai ser uma opção explorada aqui na tese. Uma vez que o objetivo central está na investigação sobre um Princípio Ponte (daqui em diante, PP) que conecte os dois estados e seus respectivos requisitos racionais, para um eliminativista, uma ponte entre tais estados é uma ponte que não leva à lugar algum.

1.3.2 Tese Lockeana

Adentrando agora especificamente as regras que visam conectar crença e probabilidade subjetiva, uma das propostas mais populares na literatura é conhecida como *Tese Lockeana* (daqui em diante, *TL*). O pioneiro dessa visão foi Richard Foley (1992). Segundo o autor, a origem desse nome ocorre em função de uma possível interpretação semelhante que já estaria presente no filósofo moderno John Locke. Nas palavras do próprio autor:

> Chamo isso de 'Tese Lockeana' não porque John Locke explicitamente a endossou — ele não fez isso — mas porque ele insinuou que a fala sobre crença é um modo geral de classificar a confiança de um indivíduo em uma proposição (FOLEY, 1992, p. 37).

Basicamente, *TL* sustenta que é racional acreditar em uma proposição somente no caso em que é racional possuir alto grau de confiança, na forma de probabilidade subjetiva, na mesma proposição. E esse "alto grau de confiança" é determinado por um ponto limitante na medida de probabilidade, isto é, um limiar s com valor maior que $1/2$ e menor que 1. Mais precisamente:

Definição 12 (Tese Lockeana)**.** $Bel(A)$ *sse* $P(A) \geq s$ (onde $1/2 < s < 1$)

Ou seja, o agente deve crer em uma proposição se, e somente se, a probabilidade subjetiva da proposição for maior ou igual que certo limiar s. *TL* é um princípio ponte na medida em que estabelece uma regra que conecta a racionalidade dos dois estados relevantes[41].

Para facilitar nosso trabalho, podemos repartir a condição de equivalência em *TL* (isto é, o "se, e somente se") em sua parte necessária e sua parte suficiente:

Definição 13 (TL Necessária)**.** Se *Bel*(A), então $P(A) \geq s$ (onde $1/2 < s < 1$)

Definição 14 (TL Suficiente)**.** Se $P(A) \geq s$, então *Bel*(A) (onde $1/2 < s < 1$)

Ou seja, a primeira condição diz que alta probabilidade subjetiva é necessária para crença, enquanto a segunda garante que alta alta probabilidade é suficiente para crença. Tal distinção nos será útil mais à frente.

Tal como foi formulada por Foley, *TL* também impõe restrições adicionais. Dado os princípios da racionalidade para crença e probabilidade subjetiva apresentados nas seções anteriores, não é racionalmente permitido acreditar em uma proposição com probabilidade subjetiva menor que 1/2, uma vez que sua negação terá probabilidade maior do que 1/2, o que conduziria o agente a acreditar em A e $\neg A$, ou seja, contradição. Chamaremos tal exigência de *Condição de Alta Probabilidade*. A exigência de que o limiar seja estabelecido abaixo de 1 está inspirado na tese bem aceita na epistemologia contemporânea de que conhecimento, justificação, e por consequência, crença, não requer certeza máxima[42]. Chamaremos tal exigência de *Condição de não-ceticismo*. Desse modo, um PP que estabeleça o valor de s como sendo máximo, ou seja $s = 1$, será um princípio cético.

Não é à toa que a *TL* é uma forma bem popular de se relacionar crença e probabilidade subjetiva. Ela parte de uma intuição bem plausível de que agentes racionais são confiantes das proposições que acreditam. Além disso, *TL* parece se adequar muito bem ao modo como nós compreendemos e utilizamos conceitos como aceitação e graus de confiança. O problema agora consiste em encontrar um valor apropriado para o limiar s, e mostrar que esse princípio se aplica consistentemente em todos os casos onde temos crença e probabilidade subjetiva. Ou seja, precisamos mostrar que no caso em que é racional possuir uma crença também é racional atribuir alta probabilidade subjetiva, na mesma proposição, acima de certo limiar s (e vice-versa).

A primeira preocupação surge da diferença entre o modelo qualitativo dos requisitos racionais para crença e o modelo quantitativo dos requisitos racionais para probabilidade subjetiva. Uma das consequências dos axiomas de Kolmogorov é que a probabilidade de uma proposição será sempre maior ou igual à probabilidade dessa

[41] Não fica claro no texto de Foley se sua proposta é "reduzir" crença à probabilidade subjetiva, de modo que a noção clássica de crença é nada mais nada menos que alta probabilidade subjetiva, ou se sua proposta é estabelecer uma conexão entre a racionalidade desses dois estados. Para os nossos propósitos, como esperado, tomarei *TL* como formulando um tipo de norma sobre a racionalidade da crença e da probabilidade subjetiva, não como um tipo de redução.

[42] Uma vertente teórica não-bayesiana pode não identificar probabilidade subjetiva máxima com certeza, mas não vamos abordar tal possibilidade aqui.

proposição em conjunção com uma outra proposição qualquer, mais precisamente: $P(A \cap B) \leq P(A)$[43]. Uma consequência disso é que, na medida em que adicionamos, por meio de conjunções, proposições com probabilidade menor que 1, a probabilidade da proposição aglomerada (ou seja, diversas proposições unidas por intersecções/conjunções) irá diminuir. Logo, a probabilidade de vários eventos independentes em conjunção diminui conforme aumentam os membros (com probabilidade estritamente menor que 1) da conjunção. Entretanto, o requisito racional 4 para crença diz que, se é racional acreditar em A e também é racional acreditar em B, minha crença em $A \cap B$ permanece racionalmente aceita, e posso expandir essa conjunção indeterminadamente enquanto creio em cada um dos conjuntos.

1.3.3 O Paradoxo da Loteria

Henry Kyburg (1961) foi um dos primeiros a sistematizar um argumento sobre a inviabilidade de TL, baseado nas considerações do parágrafo anterior. Esse argumento ficou conhecido como o *Paradoxo da Loteria*, e será apresentado resumidamente a seguir. Vamos supor um limiar s para TL, cujo valor é 0.98. Ou seja, para eu acreditar em uma determinada proposição A, minha probabilidade subjetiva em A deve ser: $P(A) \geq 0.98$. Agora considere uma loteria hipotética com 100 bilhetes, conhecida como justa, que possui exatamente um bilhete vencedor. Então é razoável, para todo bilhete A_i, com $i = 1, 2, \ldots, 100$, que se atribua uma probabilidade de 1/100 à proposição de que o bilhete A_i será premiado. Assim acreditamos de cada bilhete individual que ele irá perder, porque a probabilidade de $\neg A_i$ é de 0.99. Por exemplo, estou justificado em crer que o bilhete 03 não é o bilhete premiado, uma vez que, quando $i = 3$, $P(\neg A_3) \geq s$, logo, $Bel(\neg A_3)$ é racional em TL. Conforme a regra de aglomeração para crença, em 4, se é o caso que $Bel(\neg A_1)$ e $Bel(\neg A_2)$, então é racional acreditar na conjunção $Bel(\neg A_1 \cap \neg A_2)$, e isso expande-se para todo $\neg A_i$. Mas sabemos com certeza que, dado a loteria justa, exatamente um bilhete vencerá, isto é, $P(A_1 \cup A_2 \cup \cdots \cup A_{100}) = 1$, e $1 > s$. Portanto, acreditamos que ao menos um bilhete irá vencer, isto é, $Bel(A_1 \cup \cdots \cup A_{100})$, e também que cada bilhete individual não irá vencer, ou seja, $Bel(\neg A_1 \cap \cdots \cap \neg A_{100})$. Todavia, segundo um princípio lógico conhecido por Lei De Morgan, temos a seguinte relação de equivalência: $(A_1 \cup \cdots \cup A_{100}) = \neg(\neg A_1 \cap \cdots \cap \neg A_{100})$. Logo, o paradoxo da loteria nos mostra que TL é incompatível com os requisitos racionais para crença e probabilidade subjetiva. E isso é o caso para qualquer loteria com mais de $\frac{1}{1-s}$ bilhetes. O que o paradoxo da loteria nos mostra é que alta probabilidade subjetiva não é suficiente para crença: por maior que seja a probabilidade de $\neg A_i$, o agente pode não acreditar em

[43] A definição de probabilidade condicional nos diz que $P(A|B) = \frac{P(A \cap B)}{P(B)}$, e isso é equivalente à $P(A \cap B) = P(B|A) \cdot P(A)$. Pelos axiomas de não-negatividade e normalização, sabemos que $0 \leq P(B|A) \leq 1$, e $0 \leq P(A) \leq 1$. Com isso, a multiplicação de frações entre 0 e 1 nos garante que a igualdade $P(A \cap B) = P(B|A) \cdot P(A)$ implica a desigualdade $P(A \cap B) \leq P(A)$, chegando assim ao resultado desejado.

$\neg A_i$, sob pena de cair em contradição. Ou seja, o paradoxo da loteria é um argumento contra a parte suficiente (definição 14) de *TL*.

1.3.4 O Paradoxo do Prefácio

Uma versão semelhante de argumento contra *TL* foi formulada por David C Makinson (1965) e é conhecido como o *Paradoxo do Prefácio*. Em seu breve artigo, Makinson apresenta um contraexemplo semelhante ao seguinte caso. Suponha que um escritor de um livro científico, após a escrita de todos os capítulos, decida escrever no prefácio de sua obra, em um gesto de humildade, que seu livro contém ao menos um erro, uma alegação falsa. Ao longo de sua pesquisa e escrita, o autor afirma muitas coisas que foram descobertas, de modo que seu livro contenha um grande conjunto de afirmações que o autor julga corretas. Ele é um autor dedicado, respeitado pela academia. Assim, para cada afirmação A_i em seu livro contendo n afirmações, o autor está certo de que as proposições ali contidas são, no mínimo, altamente prováveis. Desse modo, dado *TL*, o autor acredita em cada uma de suas afirmações A_i contidas no livro e, ao mesmo tempo, deixa claro no prefácio da obra que acredita que o livro contenha ao menos um erro, ou seja, ele acredita em um conjunto de proposições como $\neg(A_1 \cap \cdots \cap A_n)$. Obviamente, este conjunto é incompatível com sua posição inicial que atribuía crença a todas proposições afirmadas no livro, isto é, no conjunto $(A_i \cap \cdots \cap A_n)$. Portanto, assim como no paradoxo da loteria, somos conduzidos, de maneira extremamente persuasiva, a aceitar um conjunto contraditório de crenças uma vez que aceitamos os requisitos pra crença e probabilidade subjetiva juntamente com *TL*. De modo semelhante, o que o paradoxo do prefácio nos mostra é que alta probabilidade subjetiva não é necessária para crença: por menor que seja a probabilidade de $(A_i \cap \cdots \cap A_n)$, o agente está autorizado a acreditar em cada uma das afirmações do livro, sob pena de cair em contradição. Ou seja, o paradoxo do prefácio é um argumento contra a parte necessária (definição 13) de *TL*.

1.3.5 Concluindo: A rota para uma resposta

Vamos resumir aqui alguns pontos do que vimos até agora. Existem dois modelos de avaliação para racionalidade de dois tipos, aparentemente distintos, de estados doxásticos: a lógica da crença como requisito racional para crenças qualitativas, e os axiomas da teoria da probabilidade (mais o princípio de condicionalização) como requisitos racionais para crenças quantitativas, isto é, probabilidades subjetivas. Apesar de reconhecer a existência de divergências, espero ter oferecido razões suficientes para convencer o leitor de que tais modelos são bem aceitos dentro de cada respectiva literatura. Contudo, ao buscar uma forma de conciliar esses dois modelos em uma estrutura uniforme, a regra mais popular, intitulada Tese Lockeana, mostrou-se claramente problemática frente aos famosos paradoxos do prefácio e da loteria. O que os paradoxos nos mostram é que alta probabilidade não é necessária, nem suficiente, para crença racional, tal como sustentado por *TL*.

Resumidamente, três pressupostos bastante plausíveis se mostraram incompatíveis frente aos contraexemplos apresentados. Assim, precisamos lidar com o seguinte conjunto de teses aparentemente contraditórias:

1. A lógica da crença: o conjunto de crenças [**B**] é consistente e fechado dedutivamente;

2. A função de probabilidade subjetiva *P* satisfaz os axiomas probabilísticos;

3. *TL* é o Princípio Ponte que governa tanto a racionalidade de [**B**] quanto de *P*.

O restante da presente Tese será sobre esse conflito. No próximo capítulo, veremos soluções a partir de mudanças em (3) e algumas saídas possíveis alterando pressupostos de (1). Em geral, não serão abordadas aqui as opções que rejeitam (2) a partir de modelos não-bayesianos para crenças quantitativas.

Meu objetivo ao longo do trabalho é mostrar que existe um requisito racional estrutural, tal como definido anteriormente, que cumpre o papel de um PP adequado. Isso significa que *TL*, por exemplo, deve ser tomado simplesmente como uma norma da racionalidade e essa norma deverá ter a forma de um requisito racional estrutural. Como consequência, se *TL* se mostrar a melhor opção de Princípio Ponte, não existe qualquer tipo de assimetria em sua relação de satisfação: frente a um desequilíbrio entre a racionalidade de estados quantitativos e qualitativos, não há qualquer hierarquia sobre qual estado deve ser mantido e qual deve ser alterado.

Algumas considerações sobre as condições de adequação do visado princípio ponte devem ser mencionadas aqui. Primeiramente, esse princípio deve ser formalmente adequado: ele não deve gerar inconsistências no sistema, e por consequência, deve dar conta de explicar os paradoxos dos tópicos anteriores. Em segundo lugar, esse princípio deve ter um grande respaldo intuitivo, de modo que o sentido de conceitos como crença, racionalidade, etc., devem se manter o mais próximo possível de seus usos técnicos e ordinários. Todavia, o que os paradoxos nos mostram é que, para preservar a exigência de consistência, alguma intuição relevante deve ser abandonada. É isso que dá o caráter de “paradoxo” aos contraexemplos apresentados. E, em terceiro lugar, é importante reforçar que esse princípio deve ser falível, ou seja, ele deve ser não-cético, no sentido de permitir crenças que não-triviais.

Antes de finalizar este primeiro capítulo, gostaria de retomar um problema deixado em aberto (de modo proposital) anteriormente. Na Seção 1.1, foi mencionado que um dos principais desafios da tese da racionalidade estrutural é lidar com o fato da simetria entre os estados em conflito. A objeção de Kolodny (2007) era exatamente de que, em uma situação de contradição entre duas crenças, posso abandonar a crença com maior respaldo evidencial e ainda ser racional, na medida que solucionei o conflito das crenças em contradição. E, na medida em que requisitos racionais estruturais não são suficientes para lidar com tais casos, a tese estruturalista possuiria consequências extremamente contra-intuitivas, e deveria, assim, ser rejeitada.

Contudo, se meu argumento sobre a existência de um Princípio Ponte na forma de um requisito racional estrutural estiver correto, o argumento de Kolodny esbarra no fato de que, ao possuir uma dupla de crenças contraditórias, além de tais crenças estarem sujeitas a um requisito estrutural de não-contradição, elas também estão sujeitas a outros requisitos racionais estruturais, e, dentre eles, está o PP em questão. Isso faz com que um requisito racional na forma de um PP seja capaz de estabelecer uma conexão entre requisitos racionais para crença (*simpliciter*) e respaldo evidencial na forma de requisitos probabilísticos. Mais precisamente, quando o agente bayesiano atualiza suas probabilidades subjetivas a partir do princípio de condicionalização, ele o faz a partir de novas evidências adquiridas (sejam elas com probabilidade subjetiva máxima, como na condicionalização estrita, ou incerta, como na condicionalização de Jeffrey). Isso faz com que, após aquisição reiterada de novas evidências ao longo do tempo, o conjunto de probabilidades subjetivas do agente bayesiano reflita o respaldo evidencial sobre suas crenças[44]. Assim, se o conjunto de probabilidades subjetivas do agente indica, em alguma medida, seu suporte evidencial (subjetivo), um requisito racional como *TL* irá garantir que crenças racionais estejam de acordo com as evidências possuídas pelo agente, uma vez que *TL* garante a ponte entre a racionalidade de crença e probabilidades subjetivas.

Aqui acredito oferecer um argumento ainda não explorado em defesa da racionalidade estrutural. Contra o defensor da racionalidade estrutural, como vimos, o substancialista argumenta que requisitos racionais estruturais, por seu caráter simétrico, permitem que o agente abandone sua crença de maior respaldo evidencial frente a uma contradição. Contudo, uma vez que tomamos *TL* como um requisito racional estrutural, frente a uma situação de contradição, o agente não poderá simplesmente rejeitar a crença na qual ele acredita possuir maior respaldo evidencial, pois essa será justamente sua crença de maior probabilidade subjetiva, uma vez que ele satisfaz *TL*. E o mais importante: isso não ocorre em função de qualquer norma substancial, ou da aceitação de requisitos de escopo estreito, mas meramente pelo fato de que requisitos racionais estruturais se aplicam como um todo, na medida em que são globais, e mais de um requisito pode ser utilizado, simultaneamente, na solução de um conflito da racionalidade. Portanto, um defensor clássico da racionalidades estrutural para crenças

[44] Sobre a relação entre suporte evidencial e condicionalização bayesiana, ver Joyce (2005). É importante reforçar aqui que o argumento apresentado acima é condicional: ele funciona apenas *se* a probabilidade subjetiva do agente refletir seu respaldo evidencial, o que um teórico substancialista ainda poderia rejeitar.

(qualitativas) tem bastante a ganhar com uma teoria da racionalidade estrutural bayesiana em conjunto com um PP apropriado, que é precisamente o que será explorado no restante do presente trabalho.

2 TEORIA DA DECISÃO COGNITIVA

> *Believe truth! Shun error! – these, we see, are two materially different laws; and by choosing between them we may end by coloring differently our whole intellectual life*.
> (William James, 1896, Seção VII)

Pense no seguinte caso hipotético. Suponha que o presidente do Banco Central do Brasil, após uma esperada reunião do COPOM (Comitê de Política Monetária) venha a público e faça a seguinte afirmação: “Nossa previsão para a taxa de inflação este ano está entre 5,3% e 5,7%”. Uma pessoa curiosa, provavelmente um filósofo, poderia se perguntar: Por que ele não afirmou que a previsão é de exatamente 5,5%? Ou que a previsão está entre 5% e 6%? Casos semelhantes são cotidianos no contexto científico. Em situações de incerteza, cientistas são conduzidos a tomarem decisões entre diferentes respostas, e muitas vezes sua escolha envolve algum tipo de *risco*. No caso apresentado, o presidente poderia ter afirmado que a taxa de inflação prevista é de 5,5%. O que acontece nesse caso? Ele escolheu transmitir a informação mais precisa possível, e aumentou assim sua chance de erro. E se ele escolher como resposta a opção entre 5% e 6%? Nesse caso, ele aumenta bastante sua chance de acertar a previsão, ao mesmo tempo que oferece uma resposta mais vaga, menos informativa, deixando muitos de seus ouvintes insatisfeitos. Fora da ciência também encontramos conflitos semelhantes em nossas práticas epistêmicas cotidianas.

O exemplo anterior apresenta uma situação onde juízos quantitativos, sobre a taxa de inflação, possuem uma margem de erro, e essa margem de erro parece refletir inclinações distintas do agente entre o risco de erro e a vontade de informar. Contudo, podemos adicionar outro elemento nesse conflito, um elemento *qualitativo*. Um exemplo típico para esse caso é, como vimos no Capítulo 1, a pergunta central do presente trabalho: dada a alta probabilidade de um evento ocorrer, quando estou autorizado a acreditar que ele de fato ocorrerá? Suponha que eu faça um teste para detecção de uma doença, e o teste dá positivo. Posso acreditar que tenho a doença sabendo que, devido a probabilidade de falso positivo no teste, a probabilidade de ter a doença é de 97%? Ou em um caso mais familiar, quando compro um único bilhete em uma (pequena) loteria de 100 números, posso acreditar que não possuo o bilhete vencedor, dada a alta probabilidade disso ocorrer (isto é, 99%)? E se a loteria contém apenas 20 números (com a probabilidade de 95% de ter o bilhete perdedor)?

No capítulo anterior, vimos que a busca por um princípio que responda alguns desses casos, o chamado Princípio Ponte (PP), nos conduz a situações incômodas de inconsistência. Neste capítulo, exploro um tipo de resposta que tenta solucionar esse conflito. Tal resposta se fundamenta principalmente sobre a ideia de que a atitude doxástica de crer envolve um tipo de "decisão cognitiva". Veremos como esse tipo de PP relaciona-se com a *Tese Lockeana* (*TL*), e pode trazer interessantes considerações sobre a natureza e valores do limiar lockeano. A primeira seção deste capítulo tem um propósito mais histórico e expositivo, e, no que fui capaz de pesquisar, apresenta a discussão com uma profundidade ímpar na literatura relevante sobre o tema. Tal retomada histórica é relevante pois, apesar do presente trabalho situar-se dentro do campo conhecido como epistemologia, ou mais precisamente, epistemologia formal (ou bayesiana), a origem filosófica da busca por um PP, e seus desafios, surgiu historicamente de problemas em filosofia da ciência. Basicamente, esse problema relevante surge de uma questão que dominava o debate entre cientistas e filósofos da ciência: como a demanda por informação influencia a legitimidade das investigações científicas? Na medida em que esforços para obter repostas verdadeiras e informativas assemelham-se a decisões em situação de incerteza, alguns autores[1] passaram a considerar um critério para inferência racional a partir de princípios gerais de tomada de decisão, aplicáveis em ambos problemas de decisões práticas e epistêmicas.

Nas seções seguintes, veremos como essa discussão aplica-se, especificamente, ao debate em torno da *racionalidade* de agentes epistêmicos, e não mais sobre a prática científica. Nesse sentido, a racionalidade é vista tal como apresentada no capítulo anterior: dado que racionalidade epistêmica, por definição, possui a verdade como alvo, ela deve nos dizer quais os melhores meios de atingi-la. É nesse contexto que uma teoria da decisão surge. Como veremos, alguns autores sugerem que tal fim epistêmico está precisamente expresso no famoso mandamento de William James, em sua obra *Will to Believe* (1956): "Maximizar a verdade e minimizar a falsidade!". O grande desafio que surge consiste no fato de que, ao buscar maximizar a verdade, o agente epistêmico naturalmente se expõe mais ao erro, e ao minimizar a falsidade, o agente torna-se cauteloso, o que tende a distanciá-lo da verdade. Desse modo, buscar a verdade é sempre uma tarefa arriscada. Assim, maximizar a verdade e minimizar a falsidade é um objetivo epistêmico que terá resultados variados relativos ao modo como ponderamos a aversão ou inclinação ao risco.

Partindo de conceitos como "maximização" e "minimização", podemos entender a escolha do agente epistêmico como uma decisão. Aqui a epistemologia se mistura com a Teoria da Decisão[2]. Uma teoria da decisão é constituída basicamente por três elementos fundamentais: (i) um objetivo a ser buscado; (ii) uma teoria da utilidade; e (iii) um modelo de tomada de decisão capaz de governar as escolhas entre diferentes opções, dependendo de como é o mundo. Não é difícil inferir que a resposta para (i)

[1] Ver, por exemplo, Levi (1967a, p.370).

[2] Teoria da Decisão é um campo de estudo sobre como agentes racionais fazem escolhas a partir de seu conjunto de crenças e desejos. Para mais detalhes sobre o tema, ver Steele e Stefánsson (2020). Uma excelente introdução sobre o tema pode ser encontrada em Peterson (2017).

está conectada à definição de racionalidade epistêmica que vimos no capítulo anterior, isto é, possuir a verdade como alvo. Assume-se que esse fim epistêmico aparece expresso no mandamento de James: Busque a verdade, evite a falsidade! Ao longo do capítulo o foco estará na busca por respostas para (ii) e (iii).

2.1 ACEITAÇÃO NO CONTEXTO CIENTÍFICO

Um dilema clássico em filosofia da ciência ocorre no conflito entre metodologias distintas para lidar com conceitos como verdade, confirmação e refutação de hipóteses. Esse conflito, de modo bem amplo, gira em torno de modelos qualitativos que buscam deduzir a verdade da conclusão a partir da verdade das premissas, e modelos quantitativos, que visam extrair informações probabilísticas, indutivas, a partir da observação reiterada de fenômenos. Filósofos da ciência como Carl Hempel (1962), entre outros de seus contemporâneos, estavam preocupados com o seguinte dilema: dado uma informação indutiva sobre um evento do mundo, em algum momento podemos passar da mera afirmação probabilística para a aceitação[3] daquele evento como verdadeiro?

No capítulo anterior, (Seção 1.3.1), foi apresentada brevemente uma vertente de filósofos da ciência e estatísticos que direcionaram sua atenção à noção de probabilidade subjetiva como uma noção substituta para aceitação e, assim, tratando o discurso sobre aceitação como uma maneira pouco rigorosa de falar sobre probabilidades subjetivas. Para Rosenkrantz (1976, p. 242), essa seria uma visão que toma a aceitação como um mero *epifenômeno*. Para os estatísticos Kendall e Stuart (1961), o uso de termos como "aceitação" ou "rejeição" de uma hipótese nas ciências é puramente convencional. O seguinte trecho ilustra bem o conflito em questão:

> Se o leitor não é capaz de superar sua aversão filosófica de tais expressões admitidamente impróprias, ele talvez irá concordar em considerá-las como palavras-código, "rejeitar" significa "decidir que as observações são desfavoráveis" e "aceitar" para o caso oposto (KENDALL; STUART, 1961, p. 163).

Como mencionado anteriormente (na Seção 1.3.1), o principal representante dessa visão na filosofia da ciência é Jeffrey (1956, 1970). Ele sustentou que deveríamos parar de falar em *aceitação racional* de uma hipótese científica, e, ao invés disso, que o papel do cientista era somente providenciar hipóteses probabilísticas que

[3] Para ser fiel ao texto dos autores, opto nesta seção pelo uso do termo "aceitação" ao invés de "crença", como feito por parte dos autores desse debate. Há uma extensa literatura filosófica sobre a natureza da crença e se esse estado corresponde ao estado de aceitação (Cf., Maher (1993)). Embora não esteja comprometido aqui com qualquer tese sobre a existência psicológica de crença e aceitação como estados distintos e irredutíveis, vou pressupor uma unicidade da racionalidade para crença em relação a racionalidade para aceitação. Portanto, assume-se que crença e/ou aceitação são simplesmente estados epistêmicos *qualitativos* sujeitos ao mesmo conjunto de requisitos racionais que foram apresentadas no capítulo anterior. Tal pressuposto assemelha-se com o que Carnap (1950) chama de *estados classificatórios*, isto é, estados que classificam uma proposição como acreditada/aceita ou não.

guiassem os agentes em sociedade. Esse ficou conhecido como um *modelo bayesiano* para inferência indutiva nas ciências. Nas palavras de Hempel (1962, p. 160), essa visão implica no abandono do modelo de aceitação, e sugere um modelo de hipóteses científicas como *tool-for-optimal-action*, ou seja, um modelo instrumentalista do conhecimento científico, onde a aceitação ou não da verdade de uma hipótese irá depender, em algum nível, da relevância prática daquela hipótese aceita.

Em seu artigo *Deductive-Nomological vs. Statistical Explanation*, Hempel (1962) foi um dos primeiros teóricos a sugerir, como uma possível alternativa a esse modelo bayesiano, uma *teoria da decisão* epistêmica para aceitação no contexto científico[4]. Alguns autores chamam isso de "Teoria da Decisão Cognitiva" (LEVI, 1967a; ROSENKRANTZ, 1976; MAHER, 1993). Esse modelo visa responder em que medida as informações contidas nas premissas de um argumento podem tornar a conclusão racional, isto é, tornar a conclusão um membro do conjunto de proposições aceitas como verdadeiras.

Um dos objetivos dos proponentes de uma teoria da decisão para a aceitação de hipóteses científicas é medir um tipo de *utilidade epistêmica*, ou seja, um valor de utilidade que não é meramente pragmático, como o defendido por Jeffrey. Para autores como Hempel (e seu contemporâneo Isaac Levi), o principal elemento que uma teoria da utilidade epistêmica deveria medir é a demanda por informação, isto é, a mediação entre o risco de acreditar erroneamente em uma proposição falsa e demanda por nova informação (como vimos nos exemplos do início do capítulo). Essa mediação deverá ser feita por um princípio de decisão racional, com uma função de utilidade epistêmica capaz de mediar tais demandas.

Voltemos à preocupação inicial de Hempel: dada uma informação indutiva sobre o mundo (por exemplo, uma sequência de lançamentos de uma moeda justa), em que momento podemos passar da mera afirmação probabilística para a aceitação daquela hipótese? Suponha que eu pretenda realizar o lançamento de uma moeda conhecidamente justa. A probabilidade de cair ao menos uma cara ao longo de dez lances é de $1-(\frac{1}{2})^{10}$, que dá um número maior que 0.999. Posso aceitar que ao menos uma cara acontecerá ao longo de dez lançamentos? A preocupação de Hempel era que tal exemplo se estendia para todas as classes de proposições indutivas nas ciências. Nas palavras do autor:

> [...] o estudo da generalização indutiva dá origem a questões sobre se é possível formular um critério para a aceitação racional de hipóteses sobre a

[4] Anteriormente à publicação de Hempel, o estatístico Neyman (1957) mencionou o fato de que a relação entre aceitação e inferência estatística envolveria conceitos decisórios. Neyman afirma que, não importa se o estudo é conduzido visando respostas práticas imediatas (como, por exemplo: qual o melhor medicamento para uma nova doença?), ou com propósitos de curiosidade científica (qual a origem do universo?), o exame de dados observacionais é seguido por um conjunto de processos mentais, e dentre eles, nas palavras do autor, "um ato de vontade, uma decisão de realizar uma ação particular" (NEYMAN, 1957, p. 14).

> base de informações que proporcionam evidência forte, porém inconclusivas, para elas (HEMPEL, 1962, p. 149).

2.1.1 Teoria da Decisão Cognitiva em Hempel

Será apresentado agora, resumidamente, como Hempel desenvolve sua teoria[5]. É prática comum assumir que o estado de crença (qualitativa) contém até três atitudes possíveis: crer, descrer, ou suspender o juízo. Em termos decisórios, essas serão as opções de "ação" para o agente em uma Teoria da Decisão Cognitiva (*TDC*). Conforme apresentado na Seção 1.2.1, $Bel(A)$ representa a atitude de crença na proposição A[6]. Agora adicionaremos dois operadores novos para as atitudes de descrença e suspensão do juízo: denotaremos a atitude de descrença por *Des* e a suspensão do juízo por *Sus*, onde $Des(A)$ significa que o agente desacredita a proposição A, e $Sus(A)$ que ele suspende o juízo sobre A[7]. Nesse sentido, $Sus(A)$ significa que o agente possui uma atitude em relação à proposição A que "está entre" a crença e a descrença[8].

Formalmente, para cada proposição $A_i \in \mathcal{A}$, o agente pode acreditar em A_i, $Bel(A_i)$; descreditar em A_i, $Des(A_i)$; ou suspender o juízo sobre A_i, $Sus(A_i)$. Assim, $\{Bel, Des, Sus\}$ é o conjunto de atitudes doxásticas disponíveis para o agente. Agora, vamos chamar de b uma função de crença que atribui para cada proposição, uma atitude doxástica, isto é, $b : \mathcal{A} \longrightarrow \{Bel, Des, Sus\}$. Apesar da notação permanecer a mesma, "*Bel*" não será agora interpretado como um operador modal de crença sobre proposições, como na lógica doxástica, mas sim como um dos elementos do conjunto de ações disponíveis ao agente em uma função b de crença (ao invés de escrever $b(A_i) = Bel$, por simplicidade, escreveremos $Bel(A_i)$). Ou seja, essas serão as três opções de "ação" (atitudes) para uma teoria da decisão cognitiva. Logo, para cada proposição pertencente a $\mathcal{A}$, o agente racional atribui racionalmente uma, e somente uma, dessas atitudes cognitivas.

---

[5] Para o bem da didática deste trabalho, farei mudanças significativas em relação às formulações originais do autor, o que nos desvia de uma interpretação totalmente fiel de sua teoria.

[6] Hempel formula sua teoria para um conjunto de *enunciados* em uma linguagem, ao invés de proposições. Reconhecidamente, proposições possuem uma estrutura menos "fina" do que sentenças em uma linguagem. No Capítulo 4 voltaremos a falar um pouco mais sobre essa distinção. Por enquanto, vamos continuar tratando proposições como os objetos da crença/aceitação, e adaptaremos a teoria de Hempel para esse caso.

[7] Algumas considerações são importante aqui. Como já dito anteriormente, não estou assumindo qualquer posição sobre a natureza psicológica desses estados mentais, se todos eles realmente existem, ou se algum é ontologicamente redutível ao outro. Minha preocupação aqui é estritamente epistêmica. Outra opção, como veremos na próxima seção com Levi (1967a), é partir de um agente altamente opinado, que possui apenas uma atitude doxástica qualitativa irredutível: a crença. Para ele, $Des(A)$ é reduzido a $Bel(\neg A)$, e suspensão do juízo sobre a proposição A seria equivalente a $Bel(A \cup \neg A)$, de modo que a única proposição mais forte aceita sobre A seria a tautologia $A \cup \neg A$, e para um agente epistêmico racional isso seria equivalente a suspensão do juízo.

[8] Aqui, "estar entre" não precisa ser tomado como uma medida quantitativa de distância. O mais correto é toma-lá como uma medida qualitativa de ranqueamento ou ordem. Na Seção 2.2.1 falo mais sobre medidas quantitativas de distância.

Estabelecidas as opções de ação para o agente epistêmico, para completar um modelo decisório é necessário definir outros dois conceitos: utilidade e consequência de uma ação. Aqui, a racionalidade de aceitar ou não uma proposição dependerá das *consequências* da respectiva ação. Em uma teoria da decisão cognitiva, cada consequência, ou resultado de uma atitude, possuirá um valor de *utilidade epistêmica*. Essa utilidade é "epistêmica", e não uma noção pragmática comum à Teoria da Decisão, ou seja, essa utilidade deverá ser portadora de um valor genuinamente epistêmico.

Agora as coisas começam a ficar mais interessantes e, por consequência, um pouco mais complexas. O valor de utilidade epistêmica será medido por meio de uma *função de utilidade*. Como esperado, o principal candidato a valor epistêmico para uma proposição acreditada é a verdade. A utilidade epistêmica será resultado da combinação da função b de estados doxásticos qualitativos, e uma função de valoração v. Para cada proposição $A_i \in \mathcal{A}$, e um mundo $w_i \in W$, temos uma função $v : \mathcal{A} \times W \longrightarrow \{0, 1\}$. Isso quer dizer que, quando A_i é verdadeira em w_i, ou seja, $w_i \in A_i$, temos que $v(A_i, w_i) = 1$, e quando A_i é falsa, ou seja, $w_i \in \neg A_i$, $v(A_i, w_i) = 0$.

Podemos agora construir uma função de utilidade u que atribui para cada par $\langle b, v \rangle$ um valor de utilidade, dependendo se uma proposição é acreditada corretamente ou não, ou seja, dependendo do resultado das funções b e v[9]. Como consequência, $u(b, v)$ possui como resultado seis combinações possíveis: (i) acreditar A quando A é verdadeira; (ii) acreditar A quando A é falsa; (iii) desacreditar A quando A é verdadeira; (iv) desacreditar A quando A é falsa; (v) suspender o juízo para A quando A é verdadeira; e (vi) suspender o juízo para A quando A é falsa.

Destas seis consequências, é bem aceito que (i) e (iv) possuem o mesmo valor epistêmico. Assim como ocorre com (ii) e (iii), e também com (v) e (vi). Logo, resta-nos somente três valores epistêmicos irredutíveis: acreditar corretamente; acreditar erroneamente; ou suspender o juízo. Assim, os possíveis resultados para a função u reduzem-se a esses três possíveis valores epistêmicos. Até então, esses resultados são bem previsíveis, dado tudo que vimos até aqui sobre a relação entre crença e o valor epistêmico da verdade. Contudo, apenas a relação com a verdade não é suficiente aqui. Para Hempel (1962, p. 154), a utilidade epistêmica de acreditar em uma proposição A não dependerá somente do valor de verdade de A, mas também do quão informativa é A.

Um novo elemento na noção de utilidade epistêmica entra em cena aqui, o valor de conteúdo informacional. Em muitos casos, ambas proposições são verdadeiras, mas ainda existe uma grande diferença entre o valor epistêmico de cada uma: enquanto uma é verdadeira, mas informa muito pouco, a outra é verdadeira e possui um rico conteúdo informativo (compare isso com o exemplo apresentado no início do capítulo

[9] Para ser mais preciso, a função u deveria possuir um subscrito relativo à proposição relevante, por exemplo u_{A_i}, uma vez que seu argumento, as funções b e v, está indexado ao mesmo elemento em $\mathcal{A}$. Por simplicidade, o subscrito será suprimido.

sobre o presidente do BC e a taxa de inflação). Esse valor epistêmico é determinado a partir de uma *medida de conteúdo* informativo da proposição.

A capacidade informativa da proposição A será medida por meio de uma função i de conteúdo informativo, que atribui para cada proposição $A_i \in \mathcal{A}$, um valor para $i(A_i)$. Hempel (1962, p. 154) apresenta as seguintes restrições para a função i:

1. $0 \leq i(A) \leq 1$;
2. $i(A) = 0$ se A for verdadeira em todos os mundos;
3. se o conteúdo de A e B for mutuamente excludente, então $i(A \cap B) = i(A) + i(B)$.

E disso o autor deriva os seguintes resultados:

4. $i(A) = 1 - i(\neg A)$;
5. se A acarreta B, então $i(A) \geq i(B)$;
6. proposições logicamente equivalente possuem a mesma medida de conteúdo.

Aqui, podemos notar uma semelhança entre os valores da função i de conteúdo e uma função de probabilidade P, tal como vimos na Seção 1.2.2. Contudo, os valores são invertidos, um é uma transformação linear do outro: quanto maior a probabilidade de uma proposição, menor seu conteúdo. Logo, $i(A)$ será igual a $1 - P(A)$, ou seja, para qualquer proposição A, $i(A)$ será equivalente à probabilidade de $\neg A$. Nesse sentido, a relação entre proposição e seu conteúdo é uma relação estritamente semântica, de modo que o conteúdo informacional diminui na medida em que aumenta a probabilidade da respectiva proposição[10].

Desse modo, Hempel (1962, p. 154) utiliza-se dos valores da função i de conteúdo informacional para determinar a utilidade epistêmica em u. Mais precisamente, temos que $u(b, v) : b \times v \longrightarrow \{i, -i, 0\}$. Isso significa que, para qualquer $A \in \mathcal{A}$, o resultado da função u será:

$$u(b, v) = \begin{cases} i(A) & \text{se } A \text{ é verdadeira e acreditada} \\ -i(A) & \text{se } A \text{ é falsa e acreditada} \\ 0 & \text{se suspende-se o juízo sobre } A. \end{cases}$$

Isso determina, finalmente, um valor de utilidade epistêmica para uma teoria da decisão cognitiva, conforme desenvolvida por Hempel. Tal utilidade epistêmica leva em conta

[10] A ideia de que hipóteses informativas são improváveis aparece nos trabalhos de Karl Popper (1959), e foi bem recebida por grande parte de seus contemporâneos. Ver Levi (1967b) e Rosenkrantz (1976, cap. 06).

dois elementos distintamente epistêmicos: a relação entre proposição acreditada e sua verdade, e o conteúdo informativo da proposição em questão.

Estabelecida uma noção de utilidade, uma Teoria da Decisão precisa determinar um modo em que o agente possa decidir da melhor forma possível sobre qual atitude tomar visando o fim desejado. Essa decisão deve levar em conta o fato de que o agente muitas vezes possui informação limitada sobre em qual mundo ele realmente está, isto é, qual $w_i \in W$ é o mundo atual. Para isso, o agente deve visar maximizar o valor epistêmico *esperado* para cada uma de suas escolhas. Dado a função u que mede a utilidade epistêmica para qualquer atitude doxástica de crença sobre qualquer elemento de $\mathcal{A}$, e uma função P de probabilidade subjetiva, um agente racional deve buscar um valor de *utilidade esperada*. Isso é feito a partir da noção de *expectativa matemática*. Dada uma partição w_i e w_j de mundos possíveis sobre W, $u_{w_i}(b, v)$ representa a utilidade epistêmica em w_i e $u_{w_j}(b, v)$ representa a utilidade epistêmica em w_j. Logo, para cada atitude em b, e cada $A \in \mathcal{A}$, podemos definir a utilidade epistêmica esperada para uma determinada atitude doxástica sobre A do seguinte modo:

$$UE(b, A) = P(A_i) \cdot u_{w_i}(b, v) + P(\neg A_i) \cdot u_{w_j}(b, v), \tag{9}$$

onde $u_{w_i}(b, v) = i(A)$ se $b = Bel$ e $w_i \in A$ (ou seja, se A for acreditada corretamente), e $u_{w_j}(b, v) = -i(A)$ se $b = Bel$ e $w_j \notin A$ (ou seja, se A for acreditada erroneamente), e $u(b, v) = 0$ se $b = Sus$, ou seja, quando A ficar em suspensão de juízo, independentemente do valor de verdade de A_i.

Estabelecido um valor de utilidade esperada, temos que, para cada possível atitude doxástica qualitativa, *Bel*, *Des*, *Sus*, sobre uma proposição A, seguem-se os seguintes resultados:

1. $UE(Bel, A) = P(A) \cdot i(A) + P(\neg A) \cdot (-i(A))$

2. $UE(Des, A) = P(\neg A) \cdot i(A) + P(A) \cdot (-i(A))$

3. $UE(Sus, A) = P(A) \cdot 0 + P(\neg A) \cdot 0 = 0$

onde $UE(Bel, A)$ representa a utilidade esperada para acreditar em A; $UE(Des, A)$ representar a utilidade esperada de desacreditar em A; e $UE(Sus, A)$ representa suspensão de juízo sobre A.

Finalmente, estabelecida uma noção de utilidade esperada capaz de ponderar valores de utilidade epistêmica e a incerteza sobre o mundo, agora precisamos do último elemento para uma teoria da decisão epistêmica: um princípio de decisão para aceitação em situações de incerteza. Hempel utilizará a famosa regra de *maximização da utilidade esperada* para guiar o agente no processo de decisão: um agente é racional se, e somente se, ele escolhe o curso de ação cujo valor de utilidade esperada é maximizado, isto é, seu valor não é excedido por nenhuma opção alternativa. Mais precisamente:

Definição 15 (*Maximização da Utilidade Esperada* - Hempel)**.** Seja b a atitude doxástica do agente em questão. Para qualquer outra atitude b^*, e uma proposição $A_i \in \mathcal{A}$,

$$b \text{ é racional } sse \ UE(b, A_i) \geq UE(b^*, A_i),$$

onde b e b^* representam, cada um, exclusivamente uma e somente uma alternativa doxástica de acreditar, desacreditar, ou suspender de juízo.

Com isso, reconstruí resumidamente uma teoria da decisão epistêmica para a formação de crença/aceitação, semelhante àquela apresentada por Hempel. Agora é possível observar algumas consequências importantes, encontradas em Hempel (1962, p. 155), que podem ser extraídas do que foi apresentado até aqui. Com um pouco de manipulação algébrica[11], encontramos uma fórmula reduzida para os resultados de utilidade esperada (1) e (2) apresentados acima:

1. $UE(Bel, A) = i(A) \cdot (2P(A) - 1)$;
2. $UE(Des, A) = i(\neg A) \cdot (1 - 2P(A))$

A partir disso, temos que a aceitação ou não de uma proposição será determinada pelo valor da probabilidade subjetiva que o agente atribui para ela, de modo que,

a) se $P(A) = \frac{1}{2}$, então os três valores possíveis para UE serão iguais a 0;
b) se $P(A) > \frac{1}{2}$, então $UE(Bel, A)$ terá valor superior as outras opções (e $UE(Sus, A)$ terá valor superior a $UE(Des, A)$); e
c) se $P(A) < \frac{1}{2}$, então $UE(Des, A)$ terá valor superior as outras (e $UE(Sus, A)$ terá valor superior a $UE(Bel, A)$).

Isso mostra que a teoria da decisão cognitiva, tal como apresentada por Hempel, curiosamente implica que aceitar A como verdadeira é racional se, e somente se, a probabilidade de A for superior a 1/2 (sim, isso realmente soa bastante Lockeano; falaremos mais sobre isso logo à frente). Contudo, Hempel reconhece a fragilidade desse resultado. Para o autor, isso ocorre em virtude da simplicidade da noção de utilidade epistêmica para aceitação, tal como foi desenvolvida (HEMPEL, 1962, p. 155). Outro problema importante já reconhecido pelo autor é o que ele chama de "não-conjuntividade da sistematização estatística": cada hipótese específica pode garantir fortemente o que está sendo afirmado, mas pode ser que tomadas individualmente elas não garantam tal força (HEMPEL, 1962, p. 163). Isso consiste basicamente nos

[11] Demonstração da primeira afirmação: $UE(Bel, A) = P(A) \cdot i(A) + P(\neg A) \cdot (-i(A)) = P(A) \cdot i(A) - P(\neg A) \cdot i(A) = P(A) \cdot i(A) - (1 - P(A)) \cdot i(A) = i(A) \cdot P(A) - 1 \cdot (1 - P(A)) = i(A) \cdot (P(A) + P(A) - 1) = i(A) \cdot (2P(A) - 1)$. A demonstração da segunda afirmação segue de modo semelhante à primeira.

paradoxos de consistência apresentados no final do Capítulo 1. Previsivelmente, esse elemento será extremamente relevante para nossa tese, e receberá a devida atenção ao longo do capítulo. Vamos agora dar mais um passo em direção à reconstrução histórica do debate, apresentado alguns elementos novos em relação a teoria inicial.

2.1.2 Teoria da Decisão Cognitiva em Levi

Uma tentativa de reformulação da teoria de Hempel, e, podemos dizer, que o primeiro desenvolvimento mais robusto de uma Teoria da Decisão Cognitiva (daqui em diante, *TDC*), apareceu em *Gambling with Truth* (1967) de Isaac Levi[12]. Como veremos, Levi desenvolveu uma teoria da decisão sobre escolhas entre hipóteses científicas concorrentes que leva em conta elementos pragmáticos e dependentes do contexto. Pretendo mostrar, brevemente, como o projeto do autor proporciona avanços relevantes em relação à teoria de Hempel, principalmente no desenvolvimento de uma nova função de utilidade epistêmica e uma nova noção de conteúdo informacional.

Em primeiro lugar, o desenvolvimento teórico de Levi parte da noção de aceitação de uma proposição[13] como resultado fruto de uma "questão relevante", denotada por $\mathcal{Q}$. Para o autor, quando nos dedicamos à ciência, visamos obter respostas para as perguntas que nos inquietam (LEVI, 1967b, p. 369). A questão relevante $\mathcal{Q}$ será uma coleção de possíveis respostas, onde cada resposta $A_i \in \mathcal{Q}$ será uma resposta completa à questão. Ainda, cada resposta em $\mathcal{Q}$ é mutuamente excludente, ou seja, para qualquer $A_i, A_j \in \mathcal{Q}$, temos que $A_i \cap A_j = \emptyset$. Esse modo de abordar a questão, faz com que a teoria de Levi seja estruturalmente diferente da abordagem de Hempel. Como veremos logo à frente, a tese de Levi diz que uma decisão cognitiva é feita sobre as diferentes escolhas possíveis de proposições, e não sobre os possíveis estados doxásticos, tal como formulado por Hempel. Logo, ao invés de visar a "melhor" atitude doxástica pra crença, Levi acredita que uma decisão cognitiva deve buscar a "melhor" resposta, dentre as opções disponíveis. Essa resposta será a menor proposição (logicamente mais forte) que contenha uma resposta para $\mathcal{Q}$. Para o autor, aceitar uma proposição é responder uma questão de modo que encontrar uma resposta é um esforço para aliviar o ceticismo.

Reforçando o que foi dito acima, o conjunto $\mathcal{Q}$ contém uma partição de potenciais respostas. Isso significa que um, e apenas um, elemento de $\mathcal{Q}$ é verdadeiro (as

[12] Novamente, não pretendo apresentar aqui uma exposição exaustiva da teoria do autor, o que levaria um trabalho exclusivo somente para isso. Meu objetivo será abordar os pontos centrais do autor no sentido do avanço que sua teoria da decisão cognitiva representa em relação à teoria de Hempel. Para isso, a presente apresentação geral será inspirada (com algumas adaptações) pelo desenvolvimento inicial do autor em *Gambling with Truth*(1967), e também por seu artigo *Information and Inference* (1967). Considerações importantes também aparecem em sua obra *The enterprise of knowledge* (1980) e em seu artigo *Acceptance revisited* (1976).

[13] Assim como em Hempel, Levi utiliza "sentenças" ao invés de proposições. Do mesmo modo, a teoria será aqui adaptada para proposições.

potenciais respostas para $\mathcal{Q}$ também formam uma *álgebra de conjuntos*, mais precisamente, um subconjunto de $\mathcal{A}$). A resposta contendo a união de todos os mundos em $\mathcal{Q}$ será representada pela proposição ($\top$), sendo uma resposta certa, mas com o menor valor informativo. A intersecção de todos os mundos em $\mathcal{Q}$ será representada pela proposição ($\bot$), representando a resposta mais informativa, porém, inconsistente (uma vez que os elementos em $\mathcal{Q}$ são mutuamente excludentes). E qualquer união de k mundos ($0 < k < n$) em $\mathcal{Q}$ será representada simplesmente por proposições A_i (com $0 < i < n$). Cada uma dessas combinações são proposições contendo potenciais respostas à questão relevante. Segundo Levi (1967b, p. 372), ($\top$) não deve ser entendido, necessariamente, como uma tautologia, uma vez que ela é capaz de conter algum, porém pequeno, conteúdo informacional, e ($\bot$) não deve ser entendido, necessariamente, como uma contradição lógica, como ficará mais claro a seguir.

Como mencionei anteriormente, a Teoria de Levi não oferece um tipo de decisão sobre qual a atitude correta frente a uma determinada proposição, como fez Hempel, mas sim, uma decisão sobre qual a proposição adequada frente à atitude de aceitação. A "proposição adequada" é entendida como a melhor resposta para $\mathcal{Q}$. Assim, a função u está definida apenas para a atitude *Bel*, ou seja, sobre qual resposta será aceita, e a decisão será agora feita sobre o conteúdo da operação $Bel(\cdot)$. Isso faz com que a teoria de Levi seja altamente opinativa: no fim, o agente deve deliberar sobre a melhor resposta para $\mathcal{Q}$, de modo que a suspensão do juízo é meramente a aceitação de ($\top$) como resposta mais forte. Ou seja, relativamente a questão relevante $\mathcal{Q}$, o agente terá uma opinião sobre todas as possíveis respostas.

Como em Hempel, a verdade ou falsidade da proposição aceita será fundamental para determinar a utilidade epistêmica da resposta tomada como correta. Assim, no mundo w em que A é verdadeira (isto é, $w \in A$), a utilidade epistêmica de $Bel(A)$ receberá um "prêmio" de utilidade. No mundo w em que A for falsa (isto é, $w \in \neg A$), a utilidade epistêmica de aceitar A como resposta receberá uma "punição" de utilidade. Seja $u(A, v)$ a utilidade de tomar uma proposição $A \in \mathcal{A}$ como resposta mais forte, onde v recebe 1 quando $v(A, w) = 1$, ou 0 se $v(A, w) = 0$. Logo, $u(A, 1)$ representa a utilidade epistêmica de aceitar A quando ela for verdadeira, e $u(A, 0)$ a utilidade epistêmica de aceitar A quando ela for falsa.

Novamente, apenas a verdade ou falsidade da proposição aceita não é suficiente para determinar sua utilidade epistêmica. Para Levi (1967a), tão importante quanto a verdade ou falsidade de uma proposição é o seu valor de conteúdo, ou seja, o quanto essa proposição é capaz de informar o agente. Como anteriormente, denotaremos tal função de conteúdo informativo por i. A função de conteúdo é responsável por oferecer o que Levi chamou de busca pelo "alívio do ceticismo": não basta que a resposta seja verdadeira, ela deve ser também informativa.

Para o autor, a medida i de conteúdo irá se relacionar com a função u de utilidade epistêmica, tal que as seguintes exigências sejam cumpridas (LEVI, 1967b, p. 377):

1. Se A é verdadeira ($w_i \in A$), para qualquer valor de $i(A)$, então $u_{w_i}(A, 1) > u_{w_i}(A, 0)$: respostas corretas possuem maior utilidade do que respostas incorretas, independente de seu conteúdo;

2. Se A e B são verdadeiras ($w_i \in A \cap B$), e $i(B) < i(A)$, então $u_{w_i}(A, 1) > u_{w_i}(B, 1)$: repostas corretas com maior conteúdo possuem maior utilidade do que respostas corretas com menor conteúdo;

3. Se é o caso que $\neg A$ e $\neg B$ ($w_i \in \neg A \cap \neg B$), e $i(B) < i(A)$, então $u_{w_i}(A, 0) > u_{w_i}(B, 0)$: respostas incorretas com maior conteúdo possuem maior utilidade do que respostas incorretas com menor conteúdo.

Para Levi, o valor informacional de aceitar uma reposta como correta é independente da resposta ser ou não correta, de modo que o valor $i(A)$ será o mesmo, independente da verdade ou falsidade de A. Destaca-se aqui a primeira diferença relevante para a função de utilidade epistêmica em Levi, relativo à teoria de Hempel: a função de utilidade em Hempel é uma transformação linear de uma medida de probabilidade, de modo que $i(A) = P(\neg A)$. Logo, em um mundo em que A é falsa, temos $u(A, 0) = -i(A)$, tal que a utilidade de acreditar erroneamente em A diminui conforme a informatividade de A aumenta. Mais precisamente, quanto maior o valor de i, menor o valor de $u(A, 0)$. Mas, segundo Levi (1967b, p. 380), o item (3) acima sustenta que a utilidade de respostas incorretas deve ser uma função decrescente em sua informatividade: respostas mais informativas são sempre preferidas a respostas menos informativas, independente do fato de ambas serem verdadeiras, ou ambas serem falsas. Mais precisamente, se $i(A) > i(B)$, então $u(A, 0) > u(B, 0)$ e $u(A, 1) > u(B, 1)$. Entretanto, para alguns autores, não está claro que essa mudança representa um ganho na teoria de Levi. De acordo com Genin (2019), em situações em que acreditamos erroneamente em algo, parece melhor aceitar proposições menos informativas do que mais informativas, uma vez que, quando em erro, seria melhor estar errado e pouco opinado do que errado e muito opinado.

Vamos agora explorar algumas consequências da formulação do autor. A partir dos itens (1)-(3) apresentados acima, Levi (1967a) demonstra que a relação entre utilidade epistêmica e conteúdo possui a seguinte forma:

- $u(A, 1) = 1 - \boldsymbol{q} \cdot i(\neg A)$
- $u(A, 0) = -\boldsymbol{q} \cdot i(\neg A)$

onde o parâmetro $\boldsymbol{q} \in [0, 1]$ é interpretado como sendo o "grau de busca por informação", representando o prêmio pelo alívio do ceticismo em oposição ao prêmio pela verdade. Tal parâmetro pondera o conflito entre evitar acreditar falsamente em algo, e a vontade de informar, aliviando, assim, o ceticismo.

Isso faz com que o parâmetro $\boldsymbol{q}$ valorize o objetivo de aceitar respostas informativas. Na medida em que $\boldsymbol{q}$ tende à 0, o valor da busca por uma resposta verdadeira predomina. Quando a resposta é correta, ou seja, $u(A, 1)$, $\boldsymbol{q} = 1$ faz com que respostas altamente informativas possuam valor de utilidade epistêmica próximo ao valor máximo, enquanto respostas pouco informativas possuam valor próximo a zero. Quando a resposta é incorreta, ou seja, $u(A, 0)$, temos que $\boldsymbol{q} = 1$ faz com que respostas pouco informativas possuam valor de utilidade próximos ao mínimo, isto é, -1, e respostas altamente informativas possuam valor de utilidade próximo a zero (aproximando-se de zero pela esquerda). Por fim, quando $\boldsymbol{q} = 0$, a utilidade de respostas corretas será máxima, enquanto a utilidade de respostas incorretas será zero, e isso ocorre independentemente do valor de conteúdo informacional da resposta em questão, pois o que se busca agora é apenas respostas corretas.

Por exemplo, suponha que A seja bastante informativa. Se aceitar que a medida de informatividade é estritamente semântica (o que Hempel aceita, mas Levi nega, como veremos à frente), temos que uma proposição informativa possui probabilidade bem baixa, de modo que a função $i(A)$ seja alta, e $i(\neg A)$ possua um valor baixo. Com isso, quando $\boldsymbol{q}$ tende à 0, $u(A, 0)$ vai de um número negativo em direção ao 0, e $u(A, 1)$ tende a 1. Isso faz com que a verdade da proposição seja o único valor epistêmico relevante para a utilidade epistêmica. Agora, quando $\boldsymbol{q}$ tende à 1, $u(A, 0)$ tende ao valor negativo do conteúdo de $i(\neg A)$, e $u(A, 1)$ tende ao valor $1 - i(\neg A)$, que é um valor alto, próximo à 1, uma vez que $i(\neg A)$ é baixo.

Por fim, uma vez que não sabemos se A é o caso ou não, precisamos ponderar os valores de $u(A, 1)$ e $u(A, 0)$ a partir do valor da probabilidade subjetiva de A em cada mundo relevante. Assim como fizemos anteriormente com a teoria de Hempel, temos agora o valor de utilidade esperada em aceitar A como resposta:

$$\begin{aligned} UE(A) =& P(A) \cdot u(A, 1) + P(\neg A) \cdot u(A, 0) \\ =& P(A) \cdot (1 - \boldsymbol{q} \cdot i(\neg A)) + (1 - P(A)) - \boldsymbol{q} \cdot i(\neg A) \\ =& P(A) - P(A) \cdot \boldsymbol{q} \cdot i(\neg A) - \boldsymbol{q} \cdot i(\neg A) + P(A) \cdot \boldsymbol{q} \cdot i(\neg A) \\ =& P(A) - \boldsymbol{q} \cdot i(\neg A) \end{aligned}$$

Assim, o parâmetro de prudência $\boldsymbol{q}$ desempenha aqui, novamente, papel central na determinação da utilidade epistêmica esperada. Quando $\boldsymbol{q}$ tende à 1, o “prêmio” por aliviar o ceticismo é máximo, e o agente deve escolher respostas menos prováveis e mais informativas. E quando $\boldsymbol{q}$ tende a 0, $UE(A)$ tende ao valor de $P(A)$, e o agente deve maximizar as respostas mais prováveis, de modo que o conteúdo informacional da resposta torna-se irrelevante frente seu risco de aceitar uma resposta falsa. Aplicando isso ao exemplo que foi apresentado na abertura do capítulo, quanto maior o valor de $\boldsymbol{q}$, mais o presidente do BC se inclina a responder que a taxa de inflação prevista é um número exato, no caso, 5,5%, e quanto menor o valor de $\boldsymbol{q}$, mais ele se inclina para uma resposta prudente e menos informativa, no caso, que a taxa de inflação prevista está entre 5% e 6%.

Por fim, resta-nos determinar um princípio decisório para que a resposta escolhida dentre as opções em $\mathcal{Q}$ maximize a utilidade epistêmica:

DEFINIÇÃO 16 (*Maximização da Utilidade Esperada* - Levi)**.** Para quaisquer duas proposições A_i e A_j candidatas a melhores respostas para $\mathcal{Q}$,

$$Bel(A_i) \text{ é racional } sse\ UE(A_i) \geq UE(A_j).$$

Estabelecido os elementos constituintes de uma *TDC* em Levi, vamos agora, brevemente, compará-la com a teoria de Hempel. Diferentemente de Hempel, Levi acredita que a racionalidade de aceitar ou não uma proposição dependerá de elementos práticos do contexto do agente, que são especificados por meio da questão relevante $\mathcal{Q}$. Para o autor, o conteúdo informacional de uma proposição específica depende, ao menos em parte, da questão para a qual tal proposição é uma resposta possível (LEVI, 1967b, p. 370). A função i de informatividade não pode ser estritamente uma relação semântica, como sustentado por Hempel, mas ela deve levar em conta aspectos pragmáticos[14]. Por isso a função i não pode simplesmente ser determinada a partir da medida de probabilidade subjetiva, ela deve levar em conta a demanda por informação para $\mathcal{Q}$, o que uma medida de probabilidade subjetiva não pode fazer, ao menos não quando probabilidade subjetiva for entendida aos moldes bayesianos (como apresentado no capítulo anterior). Na teoria de Levi (1967b), a probabilidade subjetiva $P(A)$ aparece apenas na determinação da utilidade esperada UE, e não na função de utilidade $u(A, v)$, por meio da medida de conteúdo, como acontecia na teoria de Hempel. Todavia, uma vez que tal demanda por conteúdo é fortemente dependente do contexto, nem sempre fica claro como podemos determinar precisamente o valor da função de conteúdo i.

Como consequência, se obter uma resposta correta, isto é, evitar o erro, for o único objetivo epistêmico do agente, seu valor para $\boldsymbol{q}$ deve ser máximo. Logo, o agente estaria sempre inclinado a adotar a opção mais provável, que consiste na disjunção de todas possíveis respostas, ou seja, $Bel(\top)$, o que seria equivalente à suspensão do juízo. Com isso, o agente epistêmico seria altamente cauteloso, e suspenderia o juízo

[14] Em *Gambling with Truth* (1967), Levi sugeriu que simplicidade e poder explicativo eram *desiderata* adicionais além de verdade e informatividade em problemas de decisões cognitivas. Já em *Information and Inference*(1967), o autor sugere que todos problemas cognitivos possuem apenas dois *desideratas*: verdade e informatividade. Mas agora o conceito de informatividade deve levar em conta os aspectos de simplicidade e poder explicativo a partir de uma questão $\mathcal{Q}$ relativa. Isso está refletido no modo como a função i atribui valores para os elementos de $\mathcal{Q}$.

sobre qualquer resposta, sob o risco oferecer uma resposta incorreta. Entretanto, a informatividade de sua resposta será mínima. Desse modo, a partir do valor atribuído ao parâmetro $\boldsymbol{q}$, Levi formaliza uma teoria da decisão cognitiva considerando diferentes inclinações do agente epistêmico ao risco de errar, e o objetivo de informar, e isso será de grande importância ao longo de todo o presente trabalho.

Apesar do considerável avanço teórico que a teoria da decisão cognitiva de Levi representa em relação ao projeto iniciado por Hempel, ela teve de lidar com duras objeções. Segundo Rosenkrantz (1976, p. 248), o fato de Levi se utilizar de um parâmetro $\boldsymbol{q}$ em sua teoria da utilidade epistêmica para medir tendência psicológica do agente, como sua inclinação ao risco e busca por alívio do ceticismo, tornaria a aceitação e rejeição dependente, essencialmente, de avaliações subjetivas do agente. Assim, os cientistas, individualmente, desenvolveriam suas próprias metodologias, o que poderia acarretar um tipo de "anarquismo metodológico" nas ciências.

Outra dificuldade para teoria de Levi é o fato de que ela é suscetível ao mesmo tipo relevante de desafio já detectado por Hempel, o problema da "não-conjuntividade da sistematização estatística", como vimos no final da seção anterior. No restante do capítulo tal desafio receberá uma atenção especial, agora de um ponto de vista da busca por um PP capaz de conetar a racionalidade da crença e probabilidade subjetiva a partir de uma decisão epistêmica.

2.2 TEORIA DA DECISÃO COGNITIVA COMO PRINCÍPIO PONTE

O tópico anterior apresentou as motivações e elementos iniciais presentes na origem da discussão sobre uma teoria da decisão cognitiva. Naquele contexto, a principal preocupação teórica era sobre como o cientista deveria lidar com noções como "aceitação" e "rejeição" de uma hipótese quando a informação adquirida é probabilística. Agora, durante o restante deste capítulo, pretendo mostrar como uma teoria da decisão cognitiva pode ser aplicada no debate epistemológico apresentado no capítulo anterior, mais precisamente, sobre a existência de um princípio ponte que conecte a racionalidade da crença e da probabilidade subjetiva. Questionamentos centrais se mantêm: Como passar de informações quantitativas altamente prováveis para a aceitação qualitativa de sua verdade? É alta probabilidade subjetiva *suficiente* para aceitação de uma proposição? É alta probabilidade subjetiva *necessária* para aceitação de uma proposição?

Algumas estratégias teóricas agora podem ser vistas de outra perspectiva. Enquanto o uso de ferramentas metodológicas para medir tendências psicológicas do cientista poderia soar como algo demasiadamente subversivo em Levi, aqui tais ferramentas possuem um papel, no mínimo, interessante para se avaliar algumas práticas de agentes epistêmicos e investigar a justificação e abrangência da aplicação dos requisitos racionais estruturais para crença. Veremos que o desenvolvimento teórico dessa ferramenta trouxe resultados bastante significativos para a presente discussão. Por fim, e o mais importante, a epistemologia contemporânea tem desenvolvido noções de utilidade epistêmica mais sofisticadas, capazes de impulsionar e suplementar os projetos iniciais que vimos até então.

2.2.1 Acurácia como medida de utilidade epistêmica

Começaremos esse tópico apresentando uma noção de utilidade epistêmica que nos será importante para o resto do capítulo. As teorias da decisão cognitiva desenvolvidas recentemente, e de nosso interesse prioritário, repousam sobre um conceito técnico que necessita de uma atenção especial: o conceito de *acurácia*. Para apresentar tal noção, falaremos um pouco sobre como Richard Joyce (1998) utilizou-se das ferramentas da teoria da decisão, tal como vimos acima, para motivar uma justificação puramente epistêmica para o probabilismo. Alguns dos conceitos apresentados aqui serão retomados ao longo da Tese, e desempenharão funções importantes tanto no presente capítulo quanto nos seguintes.

Como vimos no primeiro capítulo (Seção 1.2.3), Joyce (1998) é considerado um dos principais precursores da teoria da utilidade epistêmica a partir de uma medida de acurácia. Seu principal objetivo era oferecer um argumento não-pragmático a favor do probabilismo, isto é, dos axiomas da teoria da probabilidade como requisitos racionais. Joyce mostrou que qualquer distribuição de probabilidade subjetiva que não satisfaça os axiomas da teoria da probabilidade será menos acurada do que uma distribuição probabilista, independente do modo como o mundo venha a ser. Utilizando-se de uma noção decisória, Joyce mostrou que um conjunto de probabilidades subjetivas que satisfaça os axiomas da teoria da probabilidade nunca será dominado[15] por outro que não satisfaz os axiomas.

Conforme desenvolvido por Joyce (1998), para se encontrar um valor de acurácia, é necessário uma medida feita por meio de *regras de pontuação*. Uma regra de pontuação pode medir a acurácia de uma probabilidade subjetiva, ou sua *inacurácia*. No caso da medida de acurácia, quanto maior a pontuação que a medida recebe, maior sua acurácia. Previsivelmente, com uma medida de inacurácia, quanto menor a pontuação, maior a acurácia. A regra de pontuação mais popular é a *pontuação de Brier*,

[15] Em um modelo decisório, para escolher qual é certamente a melhor atitude a ser tomada, o agente precisa saber qual é o mundo atual. Todavia, na maioria dos casos, a informação que o agente possui sobre o mundo não é completa. Mas existem situações onde uma decisão será sempre mais útil, independente do modo como o mundo venha a ser, independente de qual é o mundo atual. Nesse caso, dizemos que uma escolha *domina* a outra. Como veremos, a regra decisória de dominância é uma regra mais ampla que a de maximização de utilidade esperada. Geralmente, se utiliza dominância para situações de completa ignorância (quando não há qualquer informação sobre o mundo), enquanto a maximização da utilidade esperada é utilizada em situações de risco, ou incerteza, isto é, quando nossa informação sobre o mundo é limitada, e o agente possui apenas probabilidades. Nesse sentido, toda decisão que maximiza a utilidade esperada é uma decisão não-dominada. Uma definição mais precisa de *dominância* aparecerá na Seção 2.3.3.

também conhecida como *distância euclidiana*. A pontuação de Brier diz que, para todo $A_i \in \mathcal{A}$, quando $v(A_i, w) = 1$, ou quando $v(A_i, w) = 0$, minha pontuação será resultado da distância[16] entre o valor da minha função $P(A_i)$ e sua função de valoração $v(A_i, w)$. Essa seria a medida mais elementar, também chamada de medida "local". Logo, se $v(A_i, w) = 1$ e $P(A_i) = 0.7$, minha pontuação local será $1 - 0.7 = 0.3$, onde quanto mais próximo de zero, mais próximo do alvo, e maior a acurácia. Nesse caso, temos um tipo de medida de *inacurácia*, que consiste no inverso de uma medida de acurácia: quanto menor o número, epistemicamente melhor. Generalizando, a pontuação de Briel é uma *medida* $\mathfrak{i}$ *de inacurácia global* para uma distribuição de probabilidades subjetivas, em um conjunto finito de proposições $A_1, A_2, \ldots, A_n$:

$$\mathfrak{i}(P, v) = \sum_{i=1}^{n}(v(A_i, w) - P(A_i))^2.$$

Tal modelo de justificação não-pragmática do probabilismo a partir do conceito de acurácia, com um princípio de racionalidade decisório, foi inspiração para grande parte das teorias que veremos a seguir, no restante do capítulo. Essa visão puramente epistêmica do probabilismo motivou o desenvolvimento de teorias da decisão epistêmica baseadas em uma metodologia *Accuracy First* (algo como "Primeiro Acurácia" no nosso idioma), como será apresentado a seguir.

2.2.2 *Accuracy First* na Teoria da Decisão Cognitiva

O que veremos agora consiste no desenvolvimento de uma maneira de determinar um valor de utilidade epistêmica para uma teoria da decisão cognitiva aplicando a metodologia *Accuracy First*. Isso significa que somente a acurácia, ou proximidade da verdade, de um estado deve contar para sua utilidade epistêmica.

A principal diferença das versões anteriores para a seguinte é o fato de que esta adota uma metodologia *Accuracy First*. Segundo essa visão, normas de racionalidade estão restritas às normas que promovam acurácia: racionalidade requer a adoção de métodos que espera-se que conduzam para esse fim. Segundo Kevin Dorst (2017, p. 05), um defensor da *Accuracy First*, a racionalidade deve ser um guia ótimo à verdade, e devemos começar nossa teorização *somente* visando acurácia, sem impor restrições adicionais. Nesse sentido, requisitos racionais se justificam como tais apenas na medida em que promovem acurácia. Para isso, é necessário que se desenvolva uma "medida de acurácia" para crenças, semelhante aquela desenvolvida por Joyce no tópico anterior para probabilidades subjetivas.

[16] O conceito de "distância" aqui é técnico. Trata-se basicamente do "menor caminho" (a menor medida) para ir de um ponto a outro. Em um plano, a distância entre dois pontos é medida por meio do Teorema de Pitágoras. Por isso a pontuação de Brier também é conhecida como distância euclidiana. Para mais detalhes, ver Eric Weisstein (s.d.).

O modelo decisório de fundo será semelhante aquele formulado por Hempel, na Seção 2.1.1, em que a decisão epistemicamente relevante deve ser feita sobre três atitudes possíveis: para cada $A \in \mathcal{A}$, $Bel(A)$, $Des(A)$, ou $Sus(A)$. Também serão três os valores irredutíveis assumidos aqui para as combinações de estados e valores de verdade: acreditar corretamente, acreditar erroneamente, ou suspender o juízo. Como recurso heurístico, tais resultados serão representados agora, respectivamente, pelas constantes T (do inglês *True*), F (do inglês *False*), e o valor para suspensão do juízo será neutro. T representará um "prêmio" de acurácia, enquanto F representará uma "punição". Segundo Pettigrew (2016), T e F são valores que podem ser interpretados conforme o mandamento jamesiano, como uma medida de inclinação do agente pela busca à verdade, e o risco de acreditar em uma falsidade. Agentes podem ser mais ou menos céticos ao crer em algo, e uma teoria da racionalidade da crença talvez precise levar em conta essa diversidade de tendências cognitivas.

Aqui se faz necessário chamar atenção para a diferença entre as abordagens do tópico anterior: a decisão sobre o quão informativo nosso conjunto de crenças será, não depende mais de uma medida de conteúdo, ou informatividade, específica. Como veremos mais abaixo, o valor que atribuímos para T e F será responsável por determinar a escolha do agente por uma resposta mais informativa, ou seja, que diminua o ceticismo, ou por uma resposta mais inclinada a evitar o erro, ou seja, que maximize a verdade. E tal escolha será, em acordo com a metodologia *Accuracy First*, puramente epistêmica. Nas palavras de Maher (1993, p. 233), "informação é aqui definida, em última instância, em termos das preferências cognitivas do agente". Nesse sentido, uma visão mais pragmática de conteúdo informacional, tal como defendida por Levi (1967b), é abandonada, sendo o conceito de informatividade puramente semântico e medido por meio do mandamento jamesiano apresentado acima.

Vamos agora nos aprofundar nos detalhes da construção formal de uma teoria da utilidade epistêmica por meio de uma medida de acurácia. Chamamos de *Medida de Acurácia* a função $\mathfrak{a}$ que toma um estado doxástico b, uma valoração v para uma determinada proposição $A_i \in \mathcal{A}$ em um mundo $w \in W$, e entrega uma pontuação de acurácia de b relativamente a proposição relevante (semelhante a função de utilidade em Hempel). Cada função $\mathfrak{a}$ de acurácia entrega as seguintes pontuações de utilidade epistêmica[17]:

$$\mathfrak{a}(b, v) = \begin{cases} T & \text{se } A \text{ é verdadeira e acreditada} \\ F & \text{se } A \text{ é falsa e acreditada} \\ 0 & \text{se suspende-se o juízo sobre } A \end{cases} \tag{10}$$

Aqui, T e F recebem valores contendo números reais, onde, quanto mais alta a pontuação, mais acurado é o estado. Em valores de acurácia, T é definido para $T > 0$,

[17] Sigo aqui, fundamentalmente, a apresentação de Dorst (2017).

uma vez que representa um ganho positivo, e F é definido como sempre recebendo valor negativo, tal que $F < 0$, pois representa uma perda de acurácia. Assim, temos que $F < 0 < T$. Portanto, uma medida de acurácia para crença é definida como a função $\mathfrak{a} : b \times v \longrightarrow \mathbb{R}$. Um modo alternativo de representar essa função é $\mathfrak{a}_w(Bel(A))$, cujo valor representa a acurácia de se acreditar na proposição A em um mundo w.

Agora vamos explorar um pouco mais os possíveis resultados da medida de acurácia expressa pelas variáveis T e F. Seja $|T|$ e $|F|$ o módulo de T e F[18]. É importante enfatizar que $|T|$ e $|F|$ não precisam ser iguais. Esses valores nos permite representar a distância de cada atitude para a suspensão do juízo, de modo que, quanto maior o módulo de T ou F, maior o "salto" necessário até a origem no ponto zero, onde se encontra a suspensão do juízo. Uma consequência da interpretação dos valores de $|T|$ e $|F|$ é o fato de que o agente pode ser mais ou menos conservador em suas crenças relativamente a esses valores; quando mais conservador, maior o valor de $|F|$, ou seja, errar tem um alto custo de utilidade epistêmica para o agente. Sendo assim, sem utilizar uma função específica para medir conteúdo, o defensor da *Accuracy First* é capaz de medir tendências psicológicas do agente de modo semelhante ao desenvolvido por Levi, tal como apresentado anteriormente. Essa medida de acurácia será balanceada pelo modo como entendemos o quão significativo é o valor de $|F|$ e $|T|$. Quanto maior a inclinação do agente em ter uma resposta informativa que diminua o ceticismo, maior deve ser o valor de $|T|$, e quanto maior a inclinação em evitar o erro, maior o valor de $|F|$.

Dada a noção relevante de utilidade epistêmica partindo da metodologia *Accuracy First*, agora resta apenas estabelecer os princípios decisórios para completar nossa teoria da decisão cognitiva. Novamente, os resultados mais significativos são encontrados partindo do princípio de maximização da utilidade esperada, onde a única informação disponível é uma distribuição de probabilidades subjetivas sobre o conjunto de mundos possíveis. Desse modo, a atitude racional é aquela que melhor promove $\mathfrak{a}$. Como vimos anteriormente com Hempel, o agente pode fazer isso por meio de uma estimativa, ou seja, a média ponderada da acurácia de uma atitude b em cada um dos mundos, isto é, sua *Acurácia Esperada*. Para qualquer $A_i \in \mathcal{A}$:

$$AE_{\mathfrak{a}}(b, A_i) = \sum_{w_i} P(w_i) \cdot \mathfrak{a}_{w_i}(b, v(A_i)) \qquad (11)$$

Por exemplo, suponha que um agente interessado na proposição A particione suas opções de respostas possíveis em apenas duas: A e $\neg A$, com $w_i \in A$ e $w_j \in \neg A$. O agente possui a seguinte distribuição de probabilidade subjetiva: $P(w_i) = 0.75$ e $P(w_j) = 0.25$. Agora, suponha que o agente tenha uma leve inclinação à prudência, de

[18] O módulo de um número real significa seu valor absoluto, ou seja, $|F|$ é a distância de F para a origem, que consiste no número 0. Isso faz com que o módulo de um número tenha sempre um valor positivo.

modo que seu valor para $|F|$ seja ligeiramente superior que o valor de $|T|$. No caso onde $b = Bel$, $T = 4$, $F = -5$, a acurácia esperada do agente para o estado doxástico $Bel(A)$ é de $0.75 \cdot 4 + 0.25 \cdot (-5) = 3 - 1.25 = 1.75$. Este é, portanto, o valor esperado de acurácia da atitude doxástica quando o agente possui informação limitada sobre qual é o mundo atual.

Finalmente, pra promover acurácia por conta própria, o agente deve selecionar a função b de estados doxásticos que maximiza a acurácia esperada:

DEFINIÇÃO 17 (*Maximização da Acurácia Esperada*)**.** Seja P uma função de probabilidade subjetiva e b uma função de crença, e A_i uma proposição qualquer em $\mathcal{A}$. Para qualquer $b^* \neq b$:

$$b \text{ é racional } sse\ AE_{\mathfrak{a}}(b, A_i) \geq AE_{\mathfrak{a}}(b^*, A_i).$$

Sendo assim, tem-se uma teoria da decisão cognitiva com um valor de utilidade epistêmica, exclusivamente a partir de uma medida de acurácia.

Mas o que tudo isso nos diz sobre a relação entre crença e probabilidade subjetiva? Em um primeiro momento, é fácil notar como probabilidades subjetivas aparecem na determinação de um estado de crença racional. Isso ocorre por meio na noção de acurácia esperada. Em conjunto com a metodologia *Accuracy First*, essa relação entre crença e probabilidade subjetiva parece satisfazer a condição para racionalidade epistêmica que vimos no Capítulo 1, que diz que estados doxásticos racionais devem visar a verdade. O próximo passo consiste na apresentação e análise das principais consequências de uma teoria da decisão cognitiva, com a metodologia *Accuracy First*, na busca do princípio ponte almejado.

2.3 CONSEQUÊNCIAS

2.3.1 Consequência 1: Lockeanos maximizam acurácia esperada

Suponha agora que $\mathfrak{a}$ seja determinado, como vimos, pelos valores de T e F. O que isso reflete no projeto de construção de um princípio ponte? No início do capítulo, a Seção 2.1.1 terminou com a constatação da fragilidade da *TDC* de Hempel e com um resultado bastante parecido com a famosa Tese Lockeana (*TL*) com um limiar $s = 1/2$. Ou seja, parece que existe alguma relação entre *TDC* e *TL*. Essa suposta similaridade necessita de um olhar mais atento, e é isso que será feito aqui.

Segundo Dorst (2017) e Easwaran (2016), agentes que maximizam acurácia esperada satisfazem *TL*. Isso significa que *TDC* pode ser uma forma de Princípio Ponte com uma estrutura semelhante a *TL*. E tal fato ocorre tanto com as versões iniciais de Hempel e Levi, quanto com a versão da metodologia *Accuracy First*. De fato, como ficará mais claro em seguida, qualquer teoria da decisão cognitiva que utiliza maximização da utilidade esperada como princípio decisório será equivalente a uma versão da *TL*. Mas agora é possível enriquecer *TL* com novos elementos de utilidade epistêmica capazes de tornar o valor do limiar *s* mais interessante.

Em consonância com os resultados já encontrados por Hempel (1962), Easwaran (2016) foi um dos pioneiros no desenvolvimento dos resultados que serão apresentados aqui. Por meio do cálculo de acurácia esperada, conforme a equação 11, sabemos que a acurácia esperada de $Bel(A)$ é $T \cdot P(A) + F \cdot P(\neg A)$. Uma vez que a acurácia esperada em suspender o juízo sobre *A* é 0, se *A* é o caso, então a acurácia de acreditar em *A* dever ser maior que a acurácia de suspender o juízo sobre *A*, ou seja, $Bel(A) > 0$. Assim, o estado doxástico $Bel(A)$ maximiza a acurácia esperada quando $T \cdot P(A) + F \cdot P(\neg A) > 0$ (EASWARAN, 2016, p. 829). Uma consequência importante que podemos retirar disso é que $Bel(A)$ maximiza acurácia esperada somente quando a probabilidade de *A* for maior que certa razão entre os valores de *T* e *F*. Esse resultado é generalizado por meio do seguinte teorema:

TEOREMA 2.3.1. *Seja $\mathfrak{a}$ a medida de acurácia e P uma distribuição de probabilidade subjetiva sobre W. Para cada $A \in \mathcal{A}$, e cada estado de doxástico qualitativo atribuído pela função b, temos que:*

1. *$Bel(A)$ maximiza acurácia esperada sse $P(A) \geq \frac{-F}{T-F}$*
2. *$Des(A)$ maximiza acurácia esperada sse $P(A) \leq \frac{T}{T-F}$*
3. *$Sus(A)$ maximiza acurácia esperada sse $\frac{T}{T-F} < P(A) < \frac{-F}{T-F}$*

(A demonstração desse Teorema está disponível no Apêndice B, com algumas adaptações da prova apresentada em Dorst (2017)).

Esse resultado garante que qualquer proposição acreditada possuirá uma probabilidade subjetiva correspondente maior que certo limiar, onde esse limiar será determinado por quanto o agente busca maximizar a verdade, e o quanto ele busca minimizar a falsidade. Temos assim uma nova versão de *TL*, com um limiar determinado pelo valor de *T* e *F*. Portanto, temos que *TDC* acarreta uma versão de *TL*. Conforme Dorst (2017, p. 12), o cálculo de acurácia esperada (*AE*) pode ser entendido como uma aposta epistêmica: acreditar em *A* é uma aposta que paga *T* se *A* for verdadeiro e custa *F* se *A* for falsa. Assim, uma proposição será acreditada quando sua probabilidade subjetiva for maior que a razão $\frac{-F}{T-F}$. Por exemplo, se $T = 1$ e $F = -3$, $Bel(A)$ é racional se, e somente se, $P(A) \geq 0.75$, uma vez que $\frac{-F}{T-F} = \frac{3}{4} = 0.75$. Dado essa equivalência entre *TDC* e uma forma de *TL*, as teorias apresentadas até aqui neste capítulo precisam lidar com os mesmos problemas de consistência enfrentados por *TL*. E, como vimos no início do capítulo, o próprio Hempel (1962) já reconhecia esse desafio, o assim por ele chamado “problema da não-conjuntividade da sistematização estatística”.

2.3.2 Consequência 2: Um limiar contexto (ou proposição)-dependente para o Lockeanismo

Vimos na Seção 2.2.1 que medidas de acurácia assumem o princípio de extensionalidade. Este garante que a utilidade epistêmica de um estado doxástico seja avaliada apenas a partir de dois valores: a verdade da proposição, e a atitude adotada. Nenhum outro elemento deve entrar na conta. Assim, tal pressuposto na teoria da utilidade epistêmica implica em uma versão de *TL* com um limiar fixo, relativo aos valores de T e F.

Segundo Dorst (2017, p. 188), uma teoria da utilidade epistêmica proporciona razões para esperar que atribuições de crença sejam contextualmente, e proposicionalmente, dependentes, assim como vemos na linguagem natural. Isso ocorre quando o risco de acreditar ou não em algo envolve fatores para além da simples distância da proposição alvo para com a verdade. Uma vez que o limiar para acreditar em algo é determinado pelos valores de T e F, é viável pensar que tais valores não são fixos, mas que variam conforme o contexto. Isso nos conduz ao abandono do pressuposto da extensionalidade, o que faz com que a medida de acurácia $\mathfrak{a}$ não seja uma métrica de acurácia fixa, mas sim, gerada dentro de um contexto epistêmico específico. Desse modo, para transformar *TDC* em uma versão de *TL* com limiar contextual $s = \frac{-F}{T-F}$, basta abandonar extensionalidade e aceitar que a medida de acurácia não é uma medida fixa em todas situações, com os valores de T e F fixos, mas que é gerada pelo contexto, prioridades, interesses e questões. De acordo com Dorst (2017, p. 192), essa visão toma nossas crenças como nossas melhores apostas na verdade, em resposta a nossas prioridades e contextos epistêmicos.

Outra opção, um pouco mais ousada, é sustentar que cada proposição possui um limiar de aceitação distinto em *TL*, oriundo dos valores de T e F específicos para cada proposição alvo. Agora não é apenas o contexto que pode alterar os valores de T e F, mas o conteúdo de cada proposição específica. Para isso, é necessário uma medida de acurácia $\mathfrak{a}$ que seja *proposição-dependente* (DORST, 2017, p. 189): T e F devem ser valores em funções que tomam uma proposição A de entrada, e tem como saída o valor T_A para verdades acreditadas e o desvalor F_A para falsidades acreditadas:

$$\mathfrak{a}(b, v(A)) = \begin{cases} T_A & \text{se } A \text{ é verdadeira e acreditada} \\ F_A & \text{se } A \text{ é falsa e acreditada} \\ 0 & \text{se suspende-se o juízo sobre } A \end{cases}$$

Desse modo temos uma versão de *TL* com limiar variando conforme a proposição em questão: para cada proposição A, existe um limiar $s = \frac{-F_A}{T_A-F_A}$ necessário e suficiente para crença. Essa medida de acurácia *proposição-dependente* implica no abandono da condição de *separabilidade* para a racionalidade de um conjunto de crenças. Isso significa que a acurácia do conjunto [**B**] não será mais resultado da soma da acurácia das partes, uma vez que cada proposição pode possuir um limiar próprio. Para Dorst, isso mostra que a conexão entre a teoria da utilidade epistêmica e *TL* é robusta: mesmo uma versão mínima da teoria, onde crenças são valorizadas positivamente quando verdadeiras e negativamente quando falsas, sem extensionalidade e separabilidade, ainda faz com que a maximização da acurácia esperada e a metodologia *Accuracy First* conduzam ao Lockeanismo (DORST, 2017, p. 195).

Por si só, o Lockeanismo com limiar variável não diz muito. Teorizar sobre a natureza e variabilidade do limiar s em *TL* parece a atitude mais óbvia frente aos problemas que vimos no final do primeiro capítulo. Contudo, *TDC* com a metodologia *Accuracy First* proporciona uma explicação sobre o porquê de o limiar Lockeano ser do jeito que é: ele surge do nosso fim epistêmico, da avaliação do mandamento jamesiano "acreditar na verdade e evitar falsidade". E isso pode ser fruto das prioridades contextuais do agente epistêmico.

Agora é importante dar um passo atrás e se perguntar: para onde tudo isso nos conduz? Primeiramente, ao principal problema de *TL*: os paradoxos de consistência acarretados pela aceitação simultânea dos requisitos racionais da crença, da probabilidade subjetiva, em conjunto com *TL*. Mas será que tudo que vimos até então neste capítulo traz alguma novidade sobre como devemos responder aos paradoxos de consistência?

Primeiramente, um teórico da decisão cognitiva pode assumir que o limiar $s = \frac{-F}{T-F}$ varia conforme o contexto e, em situações com loterias e prefácios, o limiar de aceitação deve ser bem maior que o limiar de grande parte dos processos ordinários de formação de crença. Por exemplo, no caso da loteria, alguém poderia alegar que o limiar para o agente acreditar na proposição "o bilhete n é um bilhete perdedor" é alto o suficiente para não permitir que o agente acredite racionalmente em tal proposição, evitando assim o paradoxo[19]. Ainda assim, isso permite que o agente continue trabalhando com limiares bem menos "céticos" em outros contextos epistêmicos.

Uma outra opção mais radical consiste em assumir que o limiar $s = \frac{-F}{T-F}$ pode variar entre proposições distintas. Isso permitiria que o agente possuísse limiares de aceitação distintos para perguntas distintas, mesmo dentro de um contexto específico. Por exemplo, o agente poderia se perguntar "O bilhete específico em minha mão é um bilhete perdedor?", de modo que o limiar de aceitação para essa resposta seja diferente do limiar da pergunta "Qualquer bilhete que eu compre é um bilhete perdedor?". E na medida em que respostas em contextos distintos não podem ser simplesmente

[19] O caso do paradoxo do prefácio é um pouco mais complicado, o que faz com que uma resposta nesse estilo seja facilmente rejeitada.

colocadas em conjunção, esse seria outra maneira de evitar o paradoxo da loteria sem abandonar a exigência de que crenças sejam consistentes e fechadas dedutivamente, ao menos dentro de um determinado contexto.

2.3.3 Consequência 3: Será o fim da consistência e fechamento dedutivo para crença?

Quando confrontados com os paradoxos de consistência, como o da loteria e do prefácio, a conclusão bastante intuitiva alcançada por grande parte dos epistemólogos poderia ser expressa no seguinte questionamento: "Não é óbvio que tais casos nos mostram que devemos abandonar ao menos algum dos requisitos racionais para crença (*simpliciter*)?" Muito autores seguiram essa linha de raciocínio. Kyburg (1961), o autor do paradoxo da loteria, desenvolveu o paradoxo justamente na tentativa de oferecer um argumento contra requisitos para crença que sempre foram tradicionalmente aceito pelos epistemólogos. Para ele, o requisito que deveria ser abandonado é o requisito de conjuntividade para crença (o requisito 4 na Seção 1.2.1): dado duas crenças racionais $Bel(A)$ e $Bel(B)$, é sempre racional acreditar na conjunção $Bel(A \cap B)$. A proposta do autor era de que, ao abandonar esse requisito de conjuntividade, seríamos capazes de manter requisitos mais importantes para crença, como o requisito de consistência e ao menos um tipo limitado de fechamento dedutivo (KYBURG, 1961, p. 56). Todavia, é bastante questionável se esse é um empreendimento viável, e alguns autores têm criticado tal estratégia[20].

A resposta mais comum aos teóricos da utilidade epistêmica é que a metologia *Accuracy First* implica no fim do requisito de consistência e/ou fechamento dedutivo para crença, tal como apresentado no capítulo 1 (Seção 1.2.1). Isso ocorre em virtude de que, para o teórico da *Accuracy First*, nenhum requisito racional deve ser aceito para além de sua contribuição na promoção de acurácia. Segundo Dorst (2017, p. 180), autores como Hempel (1962) e Levi (1967a), só não são estritamente lockeanos pois não adotam uma metodologia *Accuracy First*, e isso faz com que a exigência adicional de fechamento dedutivo, assumida por eles, entre em contradição com a versão mais simples da Tese Lockeana. Contudo, a consequência da metodologia *Accuracy First* aplicada aos requisitos racionais para crença possui interpretações distintas.

Para Dorst (2017), a teoria da decisão cognitiva em conjunto com a metodologia *Accuracy First* nos mostra que os requisitos racionais para crença são redutíveis aos requisitos probabilísticos e elementos contextuais:

> Teoria da utilidade epistêmica acarreta que — ao menos para agentes racionais — assumir crenças como atitudes irredutíveis não *faz* nada para nós. Fatos sobre crenças podem simplesmente serem 'reduzidos' a fatos sobre estados de probabilidade subjetiva e prioridades contextuais (DORST, 2017, p. 24, ênfase do autor).

[20] Ver, por exemplo, o argumento apresentado por Hannes Leitgeb (2014a).

Ou seja, a tese de Dorst diz que requisitos racionais probabilísticos aplicados à *TL* explicam toda a racionalidade da crença. Se desejamos saber qual atitude qualitativa é racional adotar em uma determinada situação, basta aplicar *TL* partindo do conjunto de probabilidades subjetivas e seus respectivos requisitos. Isso seria capaz de explicar tudo que há de racional em nossa fala sobre crença (*simpliciter*)[21].

No outro lado da moeda, Easwaran (2016) e Easwaran e Fitelson (2015) sustentam que é possível fundamentar requisitos racionais para crença a partir da teoria da utilidade epistêmica e da metodologia *Accuracy First*, mas chegam a uma conclusão oposta a Dorst[22]. Easwaran (2016) acredita que crença binária é a única atitude que deve ser considerada ao se discutir racionalidade doxástica. Nas palavras do autor:

> De modo relevante, onde bayesianos dizem que uma função de probabilidade é um estado doxástico real, e "crenças" são somente suas representações verbais, Dr. TruthLove pode dizer que *crenças* são os estados doxásticos reais, e uma função de probabilidade é somente sua representação numérica. A função de probabilidade não é tomada como representando qualquer característica real do agente — é somente um acessório para mostrar que as crenças são fortemente coerentes (EASWARAN, 2016, p. 830, ênfase do autor)

Diferentemente de Dorst, Easwaran (2016) acredita que não existe qualquer papel significativo na vida epistêmica do agente para ser desempenhado pelo que vem sendo aqui chamado de "probabilidade subjetiva". Desse modo, apenas crença importa e a justificação dos requisitos racionais para crença deve ser feita por meio da *TDC* com a metodologia *Accuracy First*. Consequentemente, *TL* deve ser tomada apenas com fins heurísticos, e não como um PP genuíno, uma vez que probabilidades aparecem somente no cálculo de utilidade esperada. Mas Easwaran (2016, p. 827) não atribui à noção de valor esperado qualquer papel relevante na racionalidade do agente. Essa é simplesmente uma ferramenta para calcular a utilidade epistêmica dos estados. A seguir, explicarei resumidamente como se fundamenta tal visão e qual sua relevância na determinação de requisitos racionais para crença.

---

[21] Um caso de redução bastante semelhante foi apresentado no capítulo anterior, Seção 1.3.1, com Christensen (2004), defendendo que, frente aos casos paradoxais de inconsistência, todos os requisitos racionais para crença devem ser abandonados, e apenas os requisitos racionais probabilísticos são suficientes para explicar o caráter aparentemente intuitivo atribuído aos tradicionais requisitos para crença. Contudo, a diferença entre Christensen (2004) e Dorst (2017) é que o primeiro não chega a fundamentar sua posição a partir de uma teoria da decisão cognitiva com a metodologia *Accuracy First*.

[22] Enquanto em Easwaran (2016) aparece claramente o objetivo de eliminar a noção de probabilidade subjetiva em defesa da prioridade da crença, em Easwaran e Fitelson (2015) o tema sobre a possibilidade de redução de um estado ao outro não é praticamente abordado. Contudo, em uma nota de rodapé, os autores afirmam que "Nossas funções de probabilidade não precisam ser funções de probabilidade subjetiva. De fato, não precisamos nem mesmo assumir aqui que agentes possuem graus de crença. Isso é importante para nós, uma vez que não queremos pressupor qualquer tipo de redução (ou eliminação) de crença cheia para (ou em favor de) crença parcial (ou qualquer outro conceito quantitativo)" (EASWARAN; FITELSON, 2015, p. 83).

Segundo Easwaran (2016), quando nossa preocupação é somente acurácia, um estado racional[23] é um estado que *não* é *dominado por acurácia*. Em Teoria da Decisão, é comum se assumir que uma atitude racional nunca é uma atitude dominada. No nosso caso, vamos assumir a seguinte exigência para o conjunto de atitudes doxásticas gerado pela função b:

DEFINIÇÃO 18 (Não-Dominância)**.** Dado uma medida de acurácia $\mathfrak{a}$, um mundo $w \in W$, uma função de crença b, e qualquer proposição $A \in \mathcal{A}$. Temos que b é uma atitude não-dominada para A se, e somente se, não existe b^*, com $b \neq b^*$, tal que:

a) Para todo $w \in W$: $\mathfrak{a}(b, A) \leq \mathfrak{a}(b^*, A)$; e

b) Existe pelo menos um $w \in W$ com: $\mathfrak{a}(b, A) < \mathfrak{a}(b^*, A)$

Como consequência, quando existir uma função b^* que seja ao menos tão acurada quanto a atitude alternativa b em todos os mundos, e mais acurada que b em pelo menos um mundo, então b é dominado por acurácia, e por consequência, b é irracional.

Estabelecido um critério de racionalidade para crença a partir de um princípio decisório com a metodologia *Accuracy First*, uma nova coleção de requisitos racionais são autorizados. Easwaran (2016) sustenta que requisitos racionais para crença baseados em acurácia não justificam a necessidade de consistência, ao menos quando falamos na consistência de um grande conjunto de proposições aceitas, o que evitaria os casos de paradoxo. Uma vez que o princípio decisório de não-dominância aplica-se para "todos os modos possíveis que o mundo venha a ter", ele é uma consequência do princípio de maximização da utilidade esperada[24]. Assim, não-dominância é um princípio mais permissível, e qualquer estado doxástico que maximize a utilidade esperada é um estado não-dominado, e por consequência, racional. Dado o resultado apresentado por meio do Teorema 2.3.1, temos uma equivalência entre maximização da utilidade esperada e *TL*, e *TL* se mostrou incompatível com fechamento dedutivo e consistência para crença. Logo, a metodologia *Accuracy First* com uma Teoria da Tecisão Cognitiva é incompatível com a nossa Lógica da crença, conforme definida no primeiro capítulo.

Embora, em Easwaran (2016), fechamento dedutivo e consistência não possam ser tomados, em sentido estrito, como requisitos racionais, isso não significa que toda contradição é racional. Como vimos anteriormente, o cálculo de utilidade esperada, e

[23] Easwaran (2016) se refere aos requisitos racionais como requisitos de coerência. Tal como definido no primeiro capítulo, requisitos racionais são requisitos estruturais, que consistem basicamente em relações de coerência. Logo, quando Easwaran (2016) fala em requisitos de coerência, vou assumir aqui que ele está falando daquilo que estou chamando de requisitos racionais estruturais.

[24] A seguinte comparação pode ser útil para compreender a relação entre o princípio de maximização da utilidade esperada e o princípio de dominância: uma atitude A é dominada por outra atitude B quando, para qualquer distribuição de probabilidade sobre W, a utilidade de A é sempre menor que a utilidade de B. Logo, atitudes que maximizam a utilidade esperada nunca são dominadas, mas atitudes não dominadas nem sempre maximizam a utilidade esperada. Para mais detalhes sobres os diferentes tipos de princípios decisórios, ver Peterson (2017).

por consequência a Tese Lockeana, depende dos valores de T e F. Relembrando, $|T|$ e $|F|$ representam respectivamente a inclinação na busca por uma resposta verdadeira e risco de acreditar em uma falsidade. Quando $|T| > |F|$, o agente torna-se "imprudente" e corre-se o risco de acreditar em uma dupla de proposições contraditórias. Isso ocorre pois, dado $|T| > |F|$, se $Bel(A) = T$ e $Bel(\neg A) = F$, então uma dupla de crenças contraditórias possui valor de acurácia $T + F > 0$, de modo que acreditar em contradições seja mais acurado que a suspensão de juízo. Portanto, quando $|T| > |F|$, acreditar em um par mutuamente inconsistente de proposições não é uma atitude rejeitada como irracional. Para evitar essa possibilidade, Easwaran (2016, p. 824) assume que $|T| < |F|$ é um ponto de partida importante para a determinação de requisitos racionais para crença.

Estabelecida a desigualdade $|T| < |F|$, um requisito mínimo para a racionalidade da crença pode ser expresso na seguinte condição de dominância:

PROPOSIÇÃO 2.3.2 (*Duplas de crenças contraditórias são sempre dominadas*)**.** *Seja* $|T| < |F|$*. Se b é uma atitude doxástica sobre uma proposição contendo uma dupla contraditória, ex.,* $Bel(A)$ *e* $Bel(\neg A)$*, então b terá uma pontuação de acurácia menor que 0 em todos os mundos, ou seja, b será dominado por uma função* b^* *que suspende o juízo sobre A.*

Isso é o caso pois, o fato de os valores de T e F serem assumidos como $|T| < |F|$ garante que uma dupla contraditória de estados doxásticos sempre receba um T e um F, para qualquer mundo possível. Como $|T| < |F|$, tal dupla contraditória terá sempre pontuação $T + F$, que sempre será negativa, dado $F < 0$. Uma vez que a suspensão do juízo recebe sempre pontuação 0, ela sempre irá dominar uma dupla de estados contraditórios.

Se duplas contraditórias são sempre dominadas, uma função b contendo uma dupla de crenças consistentes nunca é dominada. Isso ocorre pois, se b é consistente, então sempre existe um mundo onde b recebe apenas T como pontuação. Isso faz com que uma dupla de proposições consistentes nunca seja dominada por qualquer outro estado, pois b sempre receberá pontuação máxima em algum mundo.

Disso derivamos que, no mínimo, a racionalidade requer que o agente *não* acredite em uma dupla de proposições contraditórias, para qualquer que seja o caso. Na medida em que o princípio de maximização de utilidade esperada acarreta não-dominância, TL com limiar $s = \frac{-F}{T-F}$ determina o conjunto de atitudes que maximizam a utilidade esperada, consequentemente o conjunto que é não-dominado por acurácia, e por consequência, o conjunto de estados racionais. Mas agora, grandes conjuntos de crenças, como em casos de loterias e prefácios, podem conter contradições, e isso não é rejeitado como irracional pela teoria da decisão cognitiva com a metodologia *Accuracy First*. Por exemplo, pelo princípio de não-dominância, um trio inconsistente de atitudes doxásticas pode ser racional sempre que o valor de $|F|$ for menor que duas vezes o valor de T, ou seja, quando $|F| < 2|T|$.

Ao comparar os requisitos racionais para a crença que decorrem dessa visão de utilidade-como-acurácia com os requisitos apresentados no primeiro capítulo, percebemos claramente que os últimos são mais exigentes que os primeiros. Isso se reflete na aparente incapacidade dos requisitos tradicionais para crença lidarem com os paradoxos de inconsistência. A partir da fundamentação de Easwaran (2016), os novos requisitos racionais para crença são bem mais flexíveis, eles dependem dos valores de $|T|$ e $|F|$. Eles admitem até que pequenos conjuntos inconsistentes, contendo pelo menos três atitudes doxásticas, sejam admitidos como racionais.

Todavia, mesmo abrindo mão dos rígidos requisitos racionais de consistência e fechamento dedutivo, essa visão ainda admite uma grande flexibilidade para a racionalidade da crença, com requisitos mais ou menos restritivos. Isso é possível ao manipular os valores de T e F, o que implica em uma versão de *TL* com uma ampla possibilidade de limiares. Segundo Easwaran e Fitelson (2015, p. 83), tal variação do limiar poderia ser fruto de variações contextuais. Assim, o agente em questão pode maximizar a utilidade epistêmica esperada quando satisfazer uma versão de *TL* com limiar contexto-dependente, como consequência do Teorema 2.3.1 que vimos anteriormente.

Por fim, em Fitelson *et al.* (s.d., p. 37), os autores sustentam que existem contextos onde consistência e fechamento dedutivo são requisitos racionais para crença. Um desses contextos ocorre quando é exigido do conjunto de crenças que ele possua certa "harmonia", ou "unidade". Para eles, não seria razoável exigir consistência em qualquer conjunto arbitrário de crença, uma vez que tal coleção arbitrária de crenças não possui tal harmonia. Mas, por outro lado, se selecionarmos um certo conjunto de crenças que possui o tipo certo de conexão, então pode ser razoável exigir consistência de tal conjunto, naquele contexto específico, como, por exemplo, exigir que uma determinada teoria científica seja consistente, ou que o julgamento de um crime não possua proposições contraditórias.

2.4 ONDE ESTAMOS E PARA ONDE VAMOS?

No primeiro capítulo do presente trabalho, vimos que as tentativas de conectar duas noções doxásticas, que desempenham funções aparentemente distintas na literatura epistemológica, precisam lidar com um trio de suposições conflitantes: a lógica tradicional da crença, o probabilismo, e a Tese Lockeana. Agora neste capítulo foi explorada parte relevante da literatura que deu origem aos problemas apresentados no primeiro capítulo, e vimos que uma Teoria da Decisão Cognitiva (*TDC*) com a metodologia *Accuracy First* é capaz de trazer resultados interessantes para nossa busca por um PP que conecte crença e probabilidade subjetiva. A seguir farei algumas considerações sobre o que considero o melhor caminho para seguir com nossa investigação, dado o que vimos até aqui.

Resumidamente, ao longo do tópico anterior foram apresentadas duas opções para lidar com aquele trio de suposições conflitantes. Em primeiro lugar, assumir que a racionalidade da crença é dependente do contexto permite que se aceite o probabilismo juntamente com a tradicional lógica da crença (com consistência e fechamento dedutivo) e uma nova versão de *TL*, agora com valores para o limiar *s* fundamentados a partir da noção de acurácia. Em segundo lugar, é possível assumir *TL* tal como é, juntamente com o probabilismo, e rejeitar a tradicional lógica da crença, em prol de um conjunto de requisitos racionais para crença resultantes de uma decisão cognitiva visando acurácia. Esses novos requisitos para crença não exigem consistência e fechamento dedutivo, como aqueles requisitos que vimos no primeiro capítulo.

É viável afirmar que tais considerações nos colocam frente ao seguinte dilema: os paradoxos de consistência exigem que abandonemos a nossa Lógica da crença (com consistência e fechamento dedutivo), ou assumimos que a racionalidade da crença é dependente do contexto. Dificilmente se encontrará na literatura algum projeto de construção de um *PP* que esteja direcionado para outra via de solução. Antes de encerrar este capítulo, quero apresentar algumas das razões centrais pelas quais, daqui em diante, não pretendo rejeitar nossa Lógica da crença na busca por um principio ponte apropriado.

Em primeiro lugar, rejeitar consistência e fechamento dedutivo também implica em aceitar a racionalidade da crença como dependente do contexto. Argumentos a favor do abandono da consistência e fechamento dedutivo, como formulados por Easwaran (2016) e Easwaran e Fitelson (2015), dependem dos valores que são atribuídos para tendências ao risco do agente, como T e F. Isso implica que, mesmo abandonando fechamento lógico e consistência, dificilmente requisitos racionais para crença seriam todos definidos a partir de um único valor de limiar acima de $1/2$, mas que tais requisitos seriam mais ou menos restritivos conforme a necessidade. Por exemplo, como vimos, a crença em uma dupla inconsistente de proposições nunca é aceita como racional quando $|T| < |F|$. Trios inconsistentes nunca são aceitos quando $|F| < 2|T|$. Quando se expande para uma conjunção de 4 proposições mutuamente incompatíveis, tal conjunção será irracional quando $|F| < 3|T|$. E esse processo pode ser estendido indefinidamente.

Com isso, é possível afirmar que um conjunto de crenças inconsistentes (ou uma crença com um conjunto de proposições inconsistentes) poderá ser tomado como racional até quando os valores de $|T|$ e $|F|$ permitirem. Uma vez que a justificação dos requisitos racionais para crença, de qualquer modo, dependem de elementos relativos ao agente ou ao contexto, o ganho em abandonar a tradicional Lógica da crença, com aqueles requisitos apresentados no primeiro capítulo, torna-se algo bem menos significativo, na medida em que, como vimos, é possível manter consistência e fechamento dedutivo assumindo que os requisitos racionais para crença aplicam-se somente em contextos e questões específicas. Desse modo, uma versão de *TL* com limiar dependente do contexto parece ser suficiente para preservar consistência e fechamento dedutivo para crença, em conjunto com o bayesianismo. Nesse sentido, acredito ser mais atrativo, de um ponto de vista teórico, manter a consistência e fechamento dedutivo como requisitos racionais assumindo uma nova versão de *TL* com limiar flexível, do que rejeitar consistência e fechamento dedutivo para manter uma versão de *TL* com limiar fixo. Voltaremos a falar desse tema no próximo capítulo.

Em segundo lugar, até presente momento, não existe uma literatura significativa sobre requisitos diacrônicos para revisão de crença, baseados nos requisitos racionais resultantes da teoria da utilidade epistêmica por acurácia, sem consistência e fechamento dedutivo[25]. Até agora, abordou-se aqui somente a possibilidade de um PP sincrônico para a racionalidade da crença e probabilidade subjetiva. Mas o Capítulo 4 será integralmente dedicado à parte dinâmica desse debate, e veremos que existe uma diversidade de opções teóricas que visam explicar a racionalidade da atualização de crença quando uma nova informação é adquirida. Todavia, já posso adiantar ao leitor que as principais vertentes teóricas sobre esse assunto assumem que consistência e fechamento dedutivo são requisitos importantes. Sendo assim, tal empreendimento em torno da racionalidade diacrônica de uma revisão de crença ficaria bastante prejudicado sem os requisitos de consistência e fechamento dedutivo. Mas isso é assunto para o último capítulo.

Por fim, quero mencionar o que acredito ser um problema para *TDC* como um modelo de PP aos moldes do que está sendo buscado aqui. No primeiro capítulo, Seção 1.1, foi assumido que a noção de racionalidade aqui presente é uma noção estrutural, de modo que normas de racionalidades devam ser globais, de escopo amplo, simétricas, e indiferentes ao conteúdo. Mas uma teoria da decisão cognitiva como modelo de racionalidade da crença é claramente assimétrico e de escopo estreito: parte-se de um conjunto de probabilidades subjetivas e utilidades epistêmicas para, assim, determinar a racionalidade da crença. É difícil ver como uma crença racional poderia sinalizar uma falha na racionalidade do conjunto de probabilidades subjetivas. Boa parte das teorias que vimos aqui neste capítulo estão comprometidas com um tipo de prioridade da racionalidade da probabilidade subjetiva sobre a racionalidade da crença. E até mesmo no caso de Easwaran (2016), onde a noção de probabilidade é puramente heurística, a racionalidade da crença é tomada como um mero subproduto da racionalidade da decisão. E tudo isso é bastante óbvio dada a própria estrutura de uma *Teoria da Decisão*, onde existe uma hierarquia em que o conjunto de preferências e probabilidades determinam a racionalidade da ação, e nunca o contrário. No próximo capítulo, veremos como é possível utilizar vários dos resultados interessan-

[25] Acredito que o mais próximo disso é o trabalho de Shear e Fitelson (2019).

tes adquiridos até aqui para a construção de um PP não-decisório e genuinamente estrutural.

3 EM BUSCA DE UM PRINCÍPIO PONTE SINCRÔNICO

> *If a dye were marked with one figure or number of spots on four sides, and with another figure or number of spots on the two remaining sides, it would be more probable, that the former would turn up than the latter; though, if it had a thousand sides marked in the same manner, and only one side different, the probability would be much higher, and our belief or expectation of the event more steady and secure. This process of the thought or reasoning may seem trivial and obvious; but to those who consider it more narrowly, it may, perhaps, afford matter for curious speculation.*
> (David Hume, *Inquiry*, Seção VI)

Neste capítulo apresento as duas principais teorias que serão alvos de investigação e análise até o final da presente Tese. Tais teorias têm recebido destaque significativo na literatura recente sobre o tema, uma vez que oferecem versões de Princípio Ponte (PP) mantendo a exigência de fechamento dedutivo para a racionalidade da crença. A primeira é a *Teoria da Estabilidade da Crença* (LEITGEB, 2014b), ou também conhecida por *Teoria Humeana da Crença* (LEITGEB, 2015), desenvolvida por Hannes Leitgeb. A segunda, consiste na teoria *ProbaLógica* da crença (LIN; KELLY, 2012a), ou também conhecida por *Teoria do Rastreamento* (LIN; KELLY, 2012b), desenvolvida pela dupla Hanti Lin e Kevin Kelly. Ao longo do capítulo veremos como ambas teorias são fundamentadas e como elas se relacionam com a supramencionada *Tese Lockeana* (*TL*) e com a *Teoria da Decisão Cognitiva* (*TDC*) apresentada no capítulo anterior.

O principal objetivo do presente capítulo é reconstruir cuidadosamente o debate relevante e apontar virtudes e vícios em cada uma das teorias mencionadas acima. Veremos como é possível o desenvolvimento de um princípio formal que conecte crença e probabilidade subjetiva de tal modo que o resultado final seja uma teoria *consistente* e *não-cética*. Resumidamente, dizemos que um PP é consistente quando não acarreta um conjunto inconsistente de crenças para qualquer distribuição de probabilidades. E dizemos que um PP é não-cético quando ele não exige probabilidade subjetiva máxima para acreditar em uma determinada proposição. Ao longo do capítulo voltaremos a falar com mais precisão sobre tais exigências. Vamos agora a apresentação das teorias relevantes.

3.1 TEORIA DA ESTABILIDADE DA CRENÇA

No Capítulo 1, ao apresentar os requisitos iniciais para a racionalidade da crença, foi mencionada a existência de um requisito envolto de certa controvérsia, que é a exigência de "estabilidade" para a crença. Esta seção é dedicada a uma família de PPs que assumem a tese descritiva sobre a estabilidade da crença: crença é, por sua própria natureza, um estado mental estável. Mas o que seria essa estabilidade afinal? Bom, é isso que vamos explorar aqui. Nosso objetivo inicial será apresentar o que ficou conhecido por *Teoria da Estabilidade para Crença*. Inicialmente, em um espírito semelhante ao do capítulo anterior, apresento o mapeamento histórico que motivou parte significativa do debate, retomando alguns desenvolvimentos iniciais. A partir de tal retomada, podemos encontrar as primeiras tentativas explícitas de desenvolvimento de uma teoria da estabilidade que relacione crença e probabilidade subjetiva a partir da noção de "núcleo de crenças". Isso ocorreu principalmente com os trabalhos de van Fraassen (1995) e Arló-Costa (1999). Dado os devidos méritos históricos, a mais frutífera versão de uma teoria da estabilidade para crença foi desenvolvida por Hannes Leitgeb, e a maior parte do presente capítulo será dedicada a ela.

3.1.1 A teoria dos núcleos

Primeiramente, para relembrar o leitor, é importante retomar alguns pressupostos que vimos no primeiro capítulo da Tese, principalmente aqueles ligados à racionalidade da crença, Seção 1.2.1. Como vimos, nosso ponto de partida sobre a racionalidade da crença afirmar que ela é modelada por uma lógica qualitativa para estados doxásticos. Relembrando, para qualquer proposição $A \in \mathcal{A}$, temos que $Bel(A)$ satisfaz os seguintes postulados:

1. $Bel(W)$
2. não é o caso que $Bel(\emptyset)$.
3. Se $Bel(A)$ e $A \subseteq B$, então $Bel(B)$;
4. Se $Bel(A)$ e $Bel(B)$, então $Bel(A \cap B)$;

Ainda, podemos estender a quarta condição para os casos mais gerais de conjunção:

3' Para todo conjunto $\mathscr{A}$ tal que $\mathscr{A} = \{A_i \in \mathcal{A}\ ;\ Bel(A_i)\}$, é o caso que $Bel(\cap\mathscr{A})$ (Conjunção Geral).

Uma consequência desses requisitos para crença é que 3’ acarreta que, para todo conjunto $\mathcal{A}$ de proposições acreditadas, existe seu conjunto equivalente que contém apenas uma proposição, isto é, a proposição mais forte, $Bel(\cap\mathcal{A})$. O que nos interessa principalmente é quando $\mathcal{A} = [\mathbf{B}]$, ou seja, o conjunto de todas proposições acreditadas pelo agente. Nesse caso, chamemos de B_W a intersecção de todas as proposições acreditadas pelo agente (que também é uma proposição). Pelo requisito 2, B_W deve ser um subconjunto de toda proposição acreditada A_i tal que o primeiro implique logicamente o segundo. Com isso, B_W ficou conhecido como desempenhando o papel de ser um *núcleo (core) de crenças* para o agente.

Van Fraassen (1995) foi um dos primeiros autores a notar que poderia existir alguma relação frutífera entre a estrutura do núcleo de crenças e uma distribuição de probabilidades, e que isso poderia induzir alguma resposta sobre a relação entre crenças e probabilidades subjetivas. Pouco depois, Arló-Costa (1999) e Arló-Costa e Pedersen (2012), revisaram e expandiram a ideia de van Fraassen sobre a relação entre núcleos de crença e probabilidade. Chamarei aqui essa ideia geral compartilhada tanto por van Fraassen, quanto por Arló-Costa e Pederson, de “Teoria de Núcleos”.

Uma consequência imediata de um sistema de núcleos de crença é que, se $A \subseteq B$, ou seja, a verdade de A garante a verdade de B por implicação lógica, logo, $P(A) \leq P(B)$. Como toda proposição acreditada pelo agente pode ser fechada por intersecções, a intersecção de todas as proposição acreditadas pelo agente pode ser representada por uma única proposição, uma grande intersecção, conforme definido, o núcleo de crenças B_W[1]. Logo, toda proposição acreditada terá B_W como seu subconjunto, isto é, $B_W \subseteq A$, e disso segue-se que, para todo $Bel(A)$, $P(A) \geq P(B_W)$.

Ainda explorando um pouco mais as propriedades apresentadas acima, temos que núcleos de crenças cumprem uma condição que Van Fraassen (1995, p. 353) chama de *superioridade*: qualquer subconjunto não-vazio de B_W possui probabilidade significativamente maior do que qualquer subconjunto incompatível com B_W. Formalmente: para quaisquer proposições A e $B \subseteq W$, se $A \subseteq B_W$ é não-vazio, e $B \cap B_W = \emptyset$, então a probabilidade de A é significativamente maior que a probabilidade de B, ou também, A é superior à B[2]. Isso faz com que B_W seja um núcleo de crenças qualitativas para o agente, ou, nas palavras de Van Fraassen (1995, p. 349), “a intersecção que captura exatamente a imagem unificada do mundo que possui o meu consentimento”.

Na medida em que implicação lógica e probabilidade são ordenadas pela relação de subconjunto $\subseteq$, outra consequência da Teoria dos Núcleos é que qualquer

[1] Não confundir aqui a letra B presente em B_W com a letra B como uma proposição qualquer, um conjunto de mundos em W.

[2] Adaptamos a condição de van Fraassen para nossos propósitos, para garantir “probabilidade significativamente maior”, ou invés de probabilidade máxima, como prevê a teoria do autor. Arló-Costa e Pedersen (2012) mostraram que a versão de superioridade com probabilidade significativamente maior é uma instância da versão original adaptada para probabilidades monádicas. Veremos mais sobre tal distinção a seguir.

distribuição de probabilidade determinará um conjunto ordenado de núcleos[3]. Chamaremos essa família de núcleos, gerados a partir da distribuição de probabilidade P, de $\mathscr{C}_P$. Assim, a família de núcleos $\mathscr{C}_P$ é *aninhada* (do inglês, *nested*): dado quaisquer dois núcleos B_W e $B_{W'}$, ou $B_W \subseteq B_{W'}$, ou $B_{W'} \subseteq B_W$. Desse modo, dado W finito, $\mathscr{C}_P$ possui um menor núcleo, e na medida em que forma uma álgebra fechada sob uniões arbitrárias, consequentemente, contém assim um maior núcleo, ou seja, a união de todos os núcleos, que possui probabilidade 1[4]. Por fim, essa visão sugere que o núcleo mais interno, chamado de B_W, representa a proposição que compreende "tudo o que o agente acredita ser o caso", enquanto os núcleos restantes podem ser utilizados para guiar a dinâmica do conjunto de crenças do agente, governando as possíveis expansões e contrações no conjunto de crenças. Mas isso será assunto para o próximo capítulo.

Antes de continuar, aqui se faz necessário abrir um parênteses. Tanto a teoria de van Fraassen, quanto a versão de Arló-Costa, utilizam-se de uma teoria da probabilidade condicional primitiva, também conhecida como probabilidades diádicas (ou também, Probabilidade de Popper-Reni). Contrastando com a formulação clássica de Kolmogorov, a teoria da probabilidade diádica toma probabilidade condicional como primitiva, e deriva disso a noção de probabilidade absoluta. Uma das motivações para a aceitação da probabilidade diádica é que a condicionalização sobre uma proposição contraditória não é indefinida, como acontece na teoria da probabilidade clássica. Isso aconteceria quando o agente supõe uma proposição que considera impossível, isto é, com probabilidade zero. Essa propriedade desempenha uma função importante na teoria de van Fraassen, uma vez que o autor toma crenças com máxima probabilidade subjetiva não necessariamente como confiança absoluta, mas também como um tipo de *suposição* (VAN FRAASSEN, 1995, p. 351). Assim, a probabilidade diádica permite que o agente calcule a probabilidade de um evento pressupondo outro evento impossível, com probabilidade zero, sem gerar uma situação de indefinição.

Ainda, Arló-Costa e Pedersen (2012) mostraram que a teoria de núcleos pode ser formulada tanto com uma teoria da probabilidade diádica, quanto monádica, e muitas de suas propriedades e resultados são equivalentes. Assim sendo, seguindo a ideia inicial de van Fraassen, os autores concentram-se em deselvolver uma teoria dos núcleos de crença a partir de uma teoria da probabilidade diádica, o que não será feito aqui. Na medida em que no primeiro capítulo assumimos uma visão probabilista clássica (monádica), não vamos explorar os resultados e consequências do desenvol-

[3] Isso fica mais claro no exemplo que veremos na Seção 3.1.2, mais especificamente na Tabela 1. Mais detalhes podem ser encontrados em Arló-Costa e Pedersen (2012, p. 298).

[4] Se W não for finito, mas contável, na presença de *Aditividade Contável* (ver nota 25) também é possível garantir a existência de um menor núcleo, de modo que $\mathscr{C}_P$ não contenha uma cadeia infinita descendente de núcleos, isto é, $\mathscr{C}_P$ contém um menor elemento. A demonstração dessa e de outras propriedades dos núcleos de crença que apresentamos aqui podem ser encontradas em Arló-Costa (1999).

vimento de van Fraassen ou de Arló-Costa. Isso ocorre pois a noção de probabilidade diádica não cumpre os axiomas de Kolmogorov, e por hipótese assumida, não constitui um padrão de correção para probabilidades subjetivas. E é com uma concepção monádica da teoria da probabilidade que veremos um desenvolvimento recente de um PP para crença e probabilidade subjetiva emergindo da teoria dos núcleos de crença que acabamos de apresentar.

Por fim, a partir do trabalho inicial sobre a relação entre lógica e probabilidade governada por núcleos de crença, desenvolveu-se a ideia de que núcleos de crença resultariam em conjuntos de crenças *estáveis*[5]. Resumidamente, a intuição básica por trás do conceito seria a ideia de que um conjunto inicial de estados deveria ser estável no sentido de que, após refletir e fazer inferências sobre eles e a partir deles, nenhuma outra conclusão poderia ser retirada do conjunto (ARLÓ-COSTA, 1999, p. 26). Esse conceito será fundamental durante todo o presente capítulo, e principalmente no tópico seguinte. Como veremos, essa é conhecida como Teoria Humeana da crença, e foi desenvolvida por Hannes Leitgeb. A teoria de Leitgeb se assemelha bastante com o desenvolvimento de Arló-Costa [6].

3.1.2 A versão Humeana

Um dos pontos chave para compreender a motivação inicial da teoria de Leitgeb é a partir de uma mudança central na Tese Lockeana (*TL*). Leitgeb (2015, p. 152) alterou a exigência Lockeana de "probabilidade alta", pela condição de "probabilidade *estavelmente* alta". Isso constitui a essência do que o autor chamou de *Tese Humeana*[7] sobre crença (daqui em diante, *TH*). Nesse sentido, crença é entendida como um estado mental com uma probabilidade subjetiva estavelmente alta. Mas o que seria essa estabilidade? Para o autor, a estabilidade surge a partir da noção de condicionalização probabilística. Assim, uma crença estável é uma crença que não é abandonada em uma situação arbitrária de condicionalização por uma nova evidência. Veremos isso com mais detalhes mais à frente.

De modo resumido, a Tese Humeana sustenta que um agente deve acreditar

[5] A origem desse termo aparece nos trabalhos de (STALNAKER, 1993) sobre inteligência artificial.

[6] A título de curiosidade histórica, Arló-Costa e Pedersen (2012, p. 300) escrevem: "Em Junho de 2010, Hannes Leitgeb apresentou uma estimulante comunicação sobre probabilidade e regras de aceitação durante o workshop inaugural do Centro de Epistemologia Formal na Universidade de Carnegie Mellon. Durante o workshop, Leitgeb apresentou um artigo mostrando como derivar crença de graus de crença a partir de probabilidades monádicas utilizando-se da noção de estabilidade[...]. A teoria possui alguma semelhança qualitativa com a tradicional teoria dos núcleos de probabilidade, embora as definições centrais não se mostraram conectadas com uma extensão natural da teoria dos núcleos de probabilidade apresentados aqui. O objetivo da fala de Leitgeb era chegar a um acordo entre fechamento lógico e regras de aceitação com alta probabilidade". [tradução livre]

[7] De acordo com Leitgeb (2015), existe uma possível interpretação dos escritos de David Hume sobre a natureza da crença entendendo-a não somente como um "estado mental vivaz", mas também como um "estado mental estável". Novamente, os elementos históricos e interpretativos sobre esse tema fogem do escopo do presente trabalho.

(de modo *simpliciter*) em uma proposição somente no caso em que a probabilidade subjetiva do agente na proposição permanecer acima do limiar 1/2 quando condicionalizada por outra proposição que o agente considera como uma possibilidade epistêmica real[8]. Vamos deixar isso mais claro ao longo deste capítulo. De modo mais explícito, como princípio ponte, a Tese Humeana consiste no seguinte princípio:

Definição 19 ($TH_{\mathcal{Y}}$)**.** *Bel*(A) *sse* $P(A|Y) > \frac{1}{2}$, para todo Y tal que $P(Y) > 0$, e $Y \in \mathcal{Y}$

Como fizemos com a *TL*, vamos apresentar os dois sentidos do bicondicional a partir da versão necessária e da versão suficiente. Começando pela versão que afirma que probabilidade estavelmente alta é necessária para crença (isto é, não é o caso que tenho crença sem probabilidade estavelmente alta):

Definição 20 ($TH_{\mathcal{Y}}$N)**.** Se *Bel*(A), então $P(A|Y) > \frac{1}{2}$, para todo Y tal que $P(Y) > 0$, e $Y \in \mathcal{Y}$

Agora, que probabilidade estavelmente alta é suficiente para crença (isto é, não é o caso que tenho probabilidade estavelmente alta sem crença):

Definição 21 ($TH_{\mathcal{Y}}$S)**.** Se $P(A|Y) > \frac{1}{2}$, para todo Y tal que $P(Y) > 0$, e $Y \in \mathcal{Y}$, então *Bel*(A).

Como o leitor deve ter notado, apareceu um elemento *sui generis* em nossa definição: o conjunto $\mathcal{Y}$. Como mencionado anteriormente, o conjunto $\mathcal{Y}$ será interpretado como o conjunto de possibilidades epistêmicas para o agente. Mas antes de entrar especificamente nos detalhes de cada uma das condições de $TH_{\mathcal{Y}}$ (isto é, *TH* determinado pelo conjunto $\mathcal{Y}$), vamos explorar algumas possíveis interpretações para $\mathcal{Y}$, e qual sua motivação.

3.1.2.1 O conjunto $\mathcal{Y}$

Conforme visto no tópico anterior, em uma teoria de núcleos, o conjunto [**B**] é determinado por um núcleo B_W, o que ocorrerá aqui também. Ao mesmo tempo, pela definição de $TH_{\mathcal{Y}}$, [**B**] também é determinado por proposições condicionalizadas sobre elementos de $\mathcal{Y}$. Não é difícil imaginar que B_W e o conjunto $\mathcal{Y}$ de possibilidades devem possuir alguma conexão entre si, uma vez que diferentes conjuntos $\mathcal{Y}$ proporcionam diferentes conjuntos de crença, e por consequência, diferentes núcleos B_W. Como veremos, na medida em que tomamos $\mathcal{Y}$ como o conjunto de possibilidades reais para o agente, o núcleo B_W receberá uma interpretação a partir de uma semântica de mundos possíveis. Antes disso, vamos explorar um pouco as possíveis interpretações para o conjunto $\mathcal{Y}$.

[8] Inicialmente, Leitgeb parte de um limiar de 1/2 e depois expande o princípio para qualquer limiar r entre $\frac{1}{2}$ e 1. No geral, vamos assumir aqui que o limiar da Teoria Humeana é 1/2, o que será suficientes para nossos propósitos iniciais. Voltaremos a este ponto na seção 3.3.2.

Em um primeiro momento, é importante notar um aspecto epistemicamente relevante do conjunto $\mathcal{Y}$: pequenos conjuntos $\mathcal{Y}$ proporcionarão conjuntos de crenças [**B**] mais "ousados", enquanto grandes conjuntos $\mathcal{Y}$ proporcionarão conjuntos de crenças mais "prudentes". Isso parece estar em consonância com uma intuição epistêmica bem compartilhada que diz que quanto maior o número de possibilidades consideradas pelo agente, maior será a quantidade de possíveis refutadores, fazendo com que o agente seja mais cético e prudente. E, quanto mais prudente for o agente, menor será o número de proposições acreditadas por ele, isto é, menor será o conjunto [**B**]. Nesse sentido, o conjunto $\mathcal{Y}$ funciona como uma família de possíveis refutadores para as crenças do agente, isto é, proposições capazes de fazer com que o agente abandone uma de suas crenças quando considerado algum $Y \in \mathcal{Y}$.

Segundo Leitgeb (2017, p. 82), existem diversas possibilidades teóricas para os membros do conjunto $\mathcal{Y}$, e todas elas cumprem $TH_{\mathcal{Y}}$, mas algumas são mais apropriadas que outras. E uma dessas possibilidades é compreender os membros de $\mathcal{Y}$ como sendo somente as proposições cuja probabilidade é 1. Essa seria a opção mais ousada, pois o conjunto $\mathcal{Y}$ seria bem pequeno, e o agente somente poderia condicionalizar suas crenças em proposições com probabilidade 1. Essa opção seria equivalente à TH, uma vez que $P(A|Y)$, quando $Y = 1$, é equivalente a probabilidade absoluta de A, isto é, $P(A|W)$, ou simplesmente $P(A)$. O problema é que, como vimos, isso permitirá a crença em um conjunto contraditório de proposições, tal como ocorre nos paradoxos nas seções 1.3.3 e 1.3.4. Ou seja, essa interpretação para os membros de $\mathcal{Y}$ seria tão ousada que implicaria um conjunto de crenças [**B**] grande o suficiente para incluir crenças contraditórias e, portanto, B_W seria inconsistente.

Uma vez que a opção para interpretar os membros de $\mathcal{Y}$ como qualquer Y cuja probabilidade seja igual a 1 não é compatível com os requisitos lógicos para crença aqui assumidos, precisamos de opções que flexibilizem a entrada de membros para o conjunto $\mathcal{Y}$, exigindo que o nosso agente seja mais prudente: prudente o suficiente para não permitir inconsistência em B_W. Uma outra interpretação não muito eficaz, seria interpretar os membros de $\mathcal{Y}$ como qualquer Y cuja probabilidade é maior que 0. Mas, se a primeira opção era muito ousada, essa parece muito prudente, uma vez que só seria excluído do conjunto $\mathcal{Y}$ aquelas proposições cuja probabilidade é 0, ou algo muito próximo disso. Isso quer dizer que praticamente toda possibilidade lógica seria uma possibilidade epistêmica para o agente. Essa versão implicaria em um grande número de possíveis refutadores, um conjuntos $\mathcal{Y}$ muito grande, e tornaria o agente muito cético: ele teria de considerar cenários cético radicais como opções epistêmicas reais. Teoricamente, essa exigência faria com que $TH_{\mathcal{Y}}$ equivalesse crença e certeza, pois o agente só poderia acreditar em proposições com probabilidade subjetiva máxima.

Dado a inviabilidade dos dois extremos apresentados, Leitgeb (2017, p. 84) sustenta que a melhor opção para $TH_{\mathcal{Y}}$, como já adiantamos no início, é interpretar os membros de $\mathcal{Y}$ como proposições que o agente considera como possibilidades reais. "Possibilidade" aqui significa possibilidade epistêmica, não possibilidade lógica. Mais precisamente, uma proposição é uma possibilidade epistêmica real para o agente se, e somente se, o agente não acredita na falsidade dela. Ainda, denotemos por $Poss(Y)$ um operador de possibilidade sobre Y, de modo que Y é uma proposição que o agente considera uma possibilidade real, ou seja, $Y \in \mathcal{Y}$, e definimos possibilidade como sendo $Poss(Y)$:= não-$Bel(\neg Y)$[9]. Essa condição evita, ao menos em um primeiro momento, que o agente considere como possibilidade qualquer cenário cético radical, tal como demônios cartesianos e cérebros em cubas. Assim, uma crença estável será aquela que permanecer com sua probabilidade acima de 1/2 quando levamos em conta uma proposição Y, que pertence ao conjunto $\mathcal{Y}$ de possibilidades epistêmicas reais.

Formalmente, como já falamos aqui no Capítulo 1, o conjunto de crenças [**B**] deve ser interpretado a partir de uma semântica de mundos possíveis. Partindo desse pressuposto, Leitgeb (2017, p. 92) interpreta a condição de "ser uma possibilidade epistêmica real" como uma possibilidade no núcleo de crenças B_W. Com isso, uma vez que o núcleo de crenças B_W é entendido, para o agente, como a menor descrição completa de como as coisas são, ele é o conjunto de mundos doxásticamente acessíveis para o agente. Assim, dado um núcleo B_W, $Bel(A)$ é racional se, e somente se, $A \supseteq B_W$, isto é, A é verdadeira em todos os mundos pertencentes a B_W, e $Poss(A)$ é racional se, e somente se, A é verdadeira em pelo menos um mundo pertencente a B_W, ou seja, $A \cap B_W \neq \emptyset$[10].

Com isso, podemos dizer que determinada proposição Y é uma possibilidade epistêmica para o agente se, e somente se, $Y \cap B_W \neq \emptyset$. Mas deixando um pouco de lado a forma, e caminhando em direção ao conteúdo, devemos nos perguntar: Como estabelecer exatamente o que são possibilidades reais? A reposta de Leitgeb é de que a determinação final dos membros de $\mathcal{Y}$ será dependente do contexto epistêmico relevante. Cada situação irá exigir que o agente considere opções específicas como opções epistêmicas reais ou não. Nesse sentido, W contém somente as possibilidades relevantes para um determinado contexto epistêmico, visando algum propósito específico, ou decidindo sobre alguma questão específica. Tal como uma visão de mundos possíveis para proposições, o agente particiona W conforme o contexto exigir, de modo que cada célula da partição seja uma candidata real para o modo como o mundo pode ser. Tal elemento sensível ao contexto irá perpassar toda fundamentação de $TH_{\mathcal{Y}}$, e pretendo deixar isso claro nos tópicos seguintes.

---

[9] Reforçando, aqui não-$Bel(\neg Y)$ significa que $\neg Y \notin$ [**B**].

[10] No Capítulo 1, Seção 1.2.1, definimos verdade para uma proposição a partir de mundos possíveis.

3.1.2.2 *P*-estabilidade

O ponto de partida teórico para compreender $TH_{\mathcal{Y}}$ repousa sobre o conceito de estabilidade. Para Leitgeb (2017, p. 71), podemos encontrar tal conexão entre crença e estabilidade em uma possível leitura do filósofo escocês David Hume. Em suas palavras,

> Hume não apenas sustenta que estabilidade pertence à natureza da crença (disposicional) e portanto é relevante para sua filosofia da mente, mas estabilidade também suplementa a crença com uma justificação anulável [*defeasible*], e portanto é igualmente relevante para a epistemologia de Hume (LEITGEB, 2017, p. 71).

Nesse sentido, crença é entendida como uma "disposição estável para ter ideias com 'alto grau de vivacidade' sobre a qual o agir, raciocinar, e asserir, estão baseados" (LEITGEB, 2017, p.71). Esta, é basicamente a tese humeana descritiva sobre a *natureza* da crença, em contraste com a versão normativa que vimos na definição 19.

Crenças são estáveis na medida em que desempenham papel relevante no raciocínio hipotético, na asserção, e na tomada de decisão. Vamos ao seguinte exemplo. É sábado à noite, pretendo sair para comer fora com minha esposa. Decidimos que pizza é a opção da noite! Ela me questiona sobre uma opção viável que satisfaça nossos desejos. Eu respondo: "conheço uma pizzaria no centro da cidade que fica aberta até tarde, ela abre aos sábados e funciona até às 23:30!". Assim acordado, essa combinação de crenças e desejos formam meu plano de ação: tomar banho, colocar comida para o cachorro, abastecer o carro, conferir no aplicativo o trajeto mais rápido, e levar algumas moedas para dar ao flanelinha do estacionamento, caso ele esteja por lá.

Esse exemplo ilustra como seria difícil explicar situações cotidianas de asserção, raciocínio, e decisão, sem utilizar uma noção qualitativa de crença. Nesse caso, notamos que a crença em questão de que a pizzaria abre aos sábados e não fecha antes de 23:30 é *estável* sobre todas nossas condições relevantes: raciocínio, asserção, e decisão. A menos que alguma informação inesperada (pouco provável) surja durante o percurso, minha crença se mantém, de modo que eu não considere uma *possibilidade epistêmica real* que a pizzaria esteja fechada. Minha confiança nessa proposição é significativamente alta. Mas, por exemplo, eu considero uma possibilidade epistêmica que o flanelinha não esteja no estacionamento, embora eu ache mais provável que ele esteja, do que não. A intuição por trás desse caso poderia ser resumida assim: Minha crença de que a pizzaria está aberta possui uma probabilidade subjetiva alta o suficiente para não ser abandonada quando adquiro nova evidência que não seja muito pouco provável, ou seja, uma evidência inesperada para mim.

Por trás da noção de estabilidade está a ideia de que crença deve ser estável sob condicionalização probabilística[11]: quando condicionalizada por uma possibilidade

[11] A noção de estabilidade utilizada por Leitgeb aqui é inspirada pela noção de resiliência probabilística,

epistêmica do agente, a crença inicial não é abandonada. Desse modo, podemos entender a estabilidade da crença a partir de uma relação probabilística de estabilidade em um conjunto de probabilidades subjetivas. Temos assim a definição da relevante condição de *P*-estabilidade:

DEFINIÇÃO 22 (*P*-estabilidade)**.** Seja P uma distribuição de probabilidade sobre W. Para qualquer $A \subseteq W$:

$$A \text{ é } P\text{-estável } sse \ P(A \mid Y) > \tfrac{1}{2}, \text{ para todo } Y \in \mathcal{Y} \text{ tal que } A \cap Y \neq \emptyset.$$

Assim, uma proposição é *P*-estável quando sua probabilidade se mantém acima de 1/2 quando condicionalizada por qualquer proposição com a qual é compatível. Desse modo, qualquer proposição *P*-estável deve ter probabilidade maior do que sua negação. Qualquer proposição com probabilidade 1 é *P*-estável, e o conjunto vazio também é trivialmente *P*-estável (uma vez que não satisfaz $A \cap Y \neq \emptyset$). Arló-Costa e Pedersen (2012, p. 301) demonstraram que a propriedade acima é logicamente equivalente à *condição de superioridade* para núcleos de crença, conforme vimos anteriormente. Isso significa que o núcleo B_W será sempre *P*-estável, com probabilidade maior que 1/2, e, para toda proposição $A \in \mathcal{A}$ tal que $Bel(A)$, temos que $B_W \subseteq A$.

Como vimos na Definição 21 da Tese Humeana, sua parte suficiente garante que se uma determinada proposição A, condicional a uma proposição possível Y, possui probabilidade acima de 1/2, então A pertence ao conjunto de proposições acreditadas pelo agente. Isso que dizer que não é o caso que $P(A \mid Y) >$ 1/2, mas A não é acreditada pelo agente. Em outras palavras, B_W será exatamente a intersecção de todas as proposições *P*–estáveis. Agora, definida a condição de *P*-estabilidade, podemos mostrar como ela se conecta com $TH_{\mathcal{Y}}$.

Para demonstrar a relação entre a parte suficiente de $TH_{\mathcal{Y}}$ e a condição de *P*-estabilidade, tomemos A como uma proposição qualquer tal que $A \subseteq W$, e $Y \in \mathcal{Y}$, ou seja, $Poss(Y)$. Vamos supor que $P(A \mid Y) >$ 1/2, isto é, A é *P*-estável, mas que não-$Bel(A)$. Se isso gerar uma contradição, teremos $TH_{\mathcal{Y}}$S. Recordemos que, por definição, não-$Bel(A)$ é equivalente à $Poss(\neg A)$, e por consequência, $\neg A \cap B_W \neq \emptyset$. Logo, temos que $\neg A$ é uma possibilidade epistêmica para o agente, e por consequência $P(\neg A) > 0$. Se $Poss(\neg A)$, sabemos que $\neg A \in \mathcal{Y}$. Pela suposição inicial, $P(A \mid Y) >$ 1/2, mas, uma vez que $\neg A \in \mathcal{Y}$, podemos condicionalizar A por sua negação. Como $P(A \mid \neg A) = 0$, A não é *P*-estável, o que contradiz nossa hipótese inicial. Portanto, $Bel(A)$ é o caso, e temos $TH_{\mathcal{Y}}$S a partir da noção de *P*-estabilidade.

A parte $HT_{\mathcal{Y}}$N segue-se facilmente do resultado anterior. Recordemos que $HT_{\mathcal{Y}}$N garante que se uma proposição A é acreditada, então a probabilidade de A condicional a proposição possível Y é maior que 1/2. Isso que dizer que não é o caso que A pertence ao conjunto de proposições acreditadas pelo agente e $P(A \mid Y) \leq$ 1/2.

introduzida por Skyrms (1980).

Para ver como isso é, de fato, o caso, suponha $Bel(A)$. Por meio da definição de P-estabilidade (e por consequência, pelas propriedades dos núcleos de crença) sabemos que $A \supseteq B_W$, e por consequência, $P(A) \geq P(B_W)$. Seja Y uma proposição qualquer compatível com A tal que $Poss(Y)$, de modo que $P(A \mid Y) \geq P(B_W \mid Y)$. Como B_W é P-estável, e $B_W \cap Y \neq \emptyset$, então $P(B_W \mid Y) > 1/2$. Logo, $P(A \mid Y) > 1/2$, mostrando assim que A é P-estável.

Agora que estamos um pouco mais familiarizados com a noção de estabilidade, podemos, com o resultado dos parágrafos acima, apresentar então um importante resultado para a *Teoria Humeana* (LEITGEB, 2017, p. 108):

TEOREMA 3.1.1 (Teorema da representação para $TH_{\mathcal{Y}}$)**.** *Seja W finito e não-vazio, P uma distribuição de probabilidade sobre W, e r um limiar de valor* $1/2$*. As seguintes afirmações são equivalentes*[12]*:*

1. *[**B**] e P satisfazem $TH_{\mathcal{Y}}$.*
2. *Existe um conjunto não-vazio de proposições B_W, tal que:*
   a) *para todo $A \in \mathcal{A}$, $Bel(A)$ sse $B_W \subseteq A$; e*
   b) *B_W possui a seguinte propriedade de estabilidade (P-estabilidade): para todo Y com $B_W \cap Y \neq \emptyset$, e $P(Y) > 0$, é o caso que $P(B_W|Y) > 1/2$; e*
   c) *Se $P(B_W) = 1$, então B_W é a menor proposição com probabilidade 1.*

A segunda afirmação mostra que a propriedade de estabilidade da Teoria Humeana pode ser condensada em uma propriedade especial de estabilidade que se aplica somente ao núcleo de crenças no agente, ou seja, B_W. Assim, A será uma proposição acreditada somente quando $A \supseteq B_W$. Um corolário que é derivado do teorema acima é o que Leitgeb chamou de condição de Preponderância[13]:

COROLÁRIO 3.1.2 (Condição de Preponderância)**.** *Existe um conjunto não-vazio de proposições B_W tal que:*

1. *Para todo $A \in \mathcal{A}$, $Bel(A)$ sse $B_W \subseteq A$; e*
2. *B_W satisfaz o seguinte (Condição de Preponderância): Para todo $w_i \in B_W$, é o caso que $P(\{w_i\}) > P(\neg B_W)$.*

[12] O fato de que a condição de estabilidade acarreta $TH_{\mathcal{Y}}$S e $TH_{\mathcal{Y}}$N foi apresentada nos parágrafos anteriores. A demonstração da parte 1 $\rightarrow$ 2 pode ser encontrada em Leitgeb (2017, p. 108).

[13] Leitgeb demonstrou que essa condição é válida para qualquer limiar r tal que $1/2 \leq r < 1$, de modo que a versão correspondente torna-se $P(\{w_i\}) > \frac{r}{1-r} \cdot (1 - P(B_W))$. É fácil notar que quando $r = 1/2$ as duas equações são equivalentes. No final do capítulo iremos explorar as consequências de $TH_{\mathcal{Y}}$ para outros limiares acima de $1/2$.

$w_1 = A, B, C$	$w_3 = A, B, \neg C$	$w_5 = \neg A, B, \neg C$	$w_7 = \neg A, B, C$
$w_2 = A, \neg B, C$	$w_4 = A, \neg B, \neg C$	$w_6 = \neg A, \neg B, \neg C$	$w_8 = \neg A, \neg B, C$

Tabela 1 – Distribuição de possibilidades

Resumidamente, essa é uma consequência da P-estabilidade. Para ver como isso ocorre, tome A uma proposição não-vazia e P-estável, e Y uma proposição tal que $Y \cap A \neq \emptyset$, e $Y > 0$. Como vimos, a P-estabilidade garante que $P(A \mid Y) = \frac{P(A \cap Y)}{P(Y)} > r = \frac{1}{2}$, que é equivalente a $P(A \cap Y) > \frac{1}{2} \cdot P(Y)$. Pelos axiomas de Kolmogorov, e com algumas manipulações algébricas[14], descobrimos que a fórmula anterior é equivalente a $P(A \cap Y) > P(\neg A \cap Y)$. Agora, se deixarmos o valor de Y variar sobre todos seus membros de probabilidade positiva que interseccionam com A, o valor de $P(A \cap Y)$ irá variar precisamente sobre o valor dos subconjuntos de A. E de modo semelhante, o valor de $\neg A \cap Y$ irá variar sobre os subconjunto de $\neg A$. Desse modo, temos que A é P-estável se, e somente se, para qualquer $w_i \in A \cap Y$, $P(\{w_i\}) > P(\neg A)$. Segundo Leitgeb (2017, p. 192), a P-estabilidade é uma *propriedade de separação*: ela divide a classe de sub-proposições de uma dada proposição A, da classe de sub-proposições de $\neg A$, em termos de suas probabilidades (e essa é precisamente a condição de superioridade para núcleos de crença que vimos no início do capítulo).

Antes de avançar para o próximo tópico, vamos explorar um pouco mais tais resultados a partir do exemplo apresentado alguns parágrafos acima. No contexto de decisão sobre ir ou não à pizzaria, três proposições se destacam: (i) a pizzaria abre aos sábados; (ii) o flanelinha estará no estacionamento; e (iii) a pizzaria não fecha antes das 23:30. Sejam A, B, C respectivamente as proposições acima. Dadas as três proposições, nosso universo W de possibilidades consiste em 2^3 mundos possíveis, representando a atribuição de verdade ou falsidade para cada caso. A Tabela 1 representa essa distribuição entre mundos.

Entendidas como conjuntos de mundos possíveis, cada uma das proposições representará os seguintes conjuntos: $A = (\{w_1, w_2, w_3, w_4\})$; $B = (\{w_1, w_3, w_5, w_7,\})$; $C =$

[14] Demonstração:

$$\begin{aligned}
P(A \cap Y) &> 1/2 \cdot P(Y) \\
P(A \cap Y) &> 1/2 \cdot [P(A \cap Y) + P(\neg A \cap Y)] \\
P(A \cap Y) &> 1/2 \cdot P(A \cap Y) + 1/2 \cdot P(\neg A \cap Y) \\
P(A \cap Y) - 1/2 \cdot P(A \cap Y) &> 1/2 \cdot P(\neg A \cap Y) \\
P(A \cap Y) \cdot (1 - 1/2) &> 1/2 \cdot P(\neg A \cap Y) \\
P(A \cap Y) &> \frac{1/2}{1 - 1/2} \cdot P(\neg A \cap Y) \\
P(A \cap Y) &> P(\neg A \cap Y)
\end{aligned}$$

$P(\{w_1\}) = 0.4$	$P(\{w_3\}) = 0.14$	$P(\{w_5\}) = 0.02$	$P(\{w_7\}) = 0$
$P(\{w_2\}) = 0.32$	$P(\{w_4\}) = 0.11$	$P(\{w_6\}) = 0.01$	$P(\{w_8\}) = 0$

Tabela 2 – Distribuição de Probabilidade sobre W

$(\{w_1, w_2, w_6, w_7\})$. Vamos assumir aqui que o fato de a pizzaria estar aberta até às 23:30 implique que a pizzaria abre aos sábados, ou seja, $C \rightarrow A$, isto é, $\neg(C \wedge \neg A)$. Assim, dois mundos possuem probabilidade zero, w_7 e w_8. Uma possível distribuição de probabilidade sobre W é encontrada na Tabela 2.

Conforme a distribuição de probabilidade apresentada, temos que $P(A) = 0.97$, $P(B) = 0.56$ e $P(C) = 0.72$. Assim, dada a distribuição P, existem exatamente três conjuntos de crenças [**B**], tal que P e Bel satisfazem a $TH_{\mathcal{Y}}$. Esses conjuntos de crenças são gerados pela condição de estabilidade e condição de preponderância, conforme apresentado no Teorema 3.1.1 e pelo Corolário 3.1.2. Assim, nosso exemplo possui 3 opções estáveis para B_W, respectivamente: ou $B_W = \{w_1, w_2\}$ ou $B_W = \{w_1, w_2, w_3, w_4\}$, ou $B_W = \{w_1, \ldots, w_6\}$. No último caso, o agente suspende o juízo sobre qualquer resposta não-trivial para a questão relevante.

Conforme o exemplo anterior, uma das formas do conjunto de crenças do agente no exemplo cumprir $TH_{\mathcal{Y}}$ seria acreditando em $A \wedge C$, isto é, que a pizzaria abre aos sábados e não fecha antes das 23:30. Esta opção de B_W seria $\{w_1, w_2\}$, e nesse caso o agente consideraria B como uma possibilidade real. Também pode ser o caso em que o agente seja mais cauteloso e satisfaça $TH_{\mathcal{Y}}$ acreditando somente que a pizzaria abre aos sábados, e deixando em aberto se ela fecha ou não antes das 23:30. Nesse caso, B_W seria $\{w_1, w_2, w_3, w_4\}$. Ou ainda poderia ser o caso que o agente suspende o juízo sobre as três proposições e toma B_W como sendo $\{w_1, \ldots, w_6\}$.

Em cada um dos casos, o conjunto de crenças [**B**] é estável. Por exemplo, no caso em que $B_W = \{w_1, w_2\}$, o agente acredita que $(A \wedge C)$ é o caso, e $\neg B$ é uma possibilidade para o agente, uma vez que não é o caso que $Bel(B)$. Segundo nosso exemplo, $P(\neg B) = 0.44$. Portanto, a crença de que a pizzaria está aberta e não fecha antes das 23:30 é estável quando condicionalizada sobre a possibilidade do flanelinha não estar no estacionamento:

$$P(A \cap C|\neg B) = P(B_W|\neg B) = \frac{P(B_W \cap \neg B)}{P(\neg B)} = \frac{0.32}{0.44} = 0.72 > r = \frac{1}{2}.$$

Isso é exatamente uma das condições de $HT_{\mathcal{Y}}$ assumindo que crenças são estáveis.

3.1.2.3 O Paradoxo da Loteria novamente

Uma consequência para $HT_{\mathcal{Y}}$ que podemos extrair a partir do que vimos até então é que crenças são fortemente sensíveis ao contexto. Formalmente, o agente irá modelar o contexto por meio do modo como ele "particiona" W. Segundo Leitgeb (2015,

p. 145), esse aspecto de $TH_{\mathcal{Y}}$ ilumina o principal desafio que enfrentamos até agora na construção de um princípio ponte: o paradoxo da loteria. Suponha o seguinte contexto: o agente está interessado em qual bilhete será o vencedor. No caso com $A_1, \ldots, A_{100}$ bilhetes, o agente busca a seguinte resposta: o bilhete A_i será ou não sorteado? Em uma loteria de 100 bilhetes, temos 100 resultados possíveis, ou seja, 100 mundos possíveis: $W = \{w_1; \ldots; w_{100}\}$. Nesse contexto, particionamos o universo de probabilidades em 100 mundos igualmente possíveis e prováveis: $\{\{w_1\}; \{w_2\}; \ldots; \{w_{100}\}\}$ onde cada $\{w_i\}$ é uma possibilidade considerada pelo agente, com $P(\{w_i\}) = \frac{1}{100}$ para cada bilhete A_i. Nesse caso, o único conjunto B_W estável é propriamente o universo W, ou seja, a proposição de probabilidade máxima "algum bilhete irá vencer". Assim, nosso agente acredita que algum bilhete será sorteado, mas ele não acredita na derrota de qualquer um dos bilhetes individualmente, ou seja, ele não acredita que o bilhete A_i não será sorteado (para qualquer A_i), evitando assim, a inconsistência. Esse seria um caso extremo, onde o conjunto de possibilidades naquele contexto conduziria o agente a uma posição extremamente prudente, onde crença se aproxima de certeza, pois o agente terminaria acreditando somente na única proposição com probabilidade 1.

Agora suponha um outro contexto no qual o agente quer saber se um bilhete específico será sorteado ou não. Isso pode ocorrer se, ao comprar o bilhete de determinado número, o agente está preocupado se deve ou não acreditar na derrota de um bilhete específico, por exemplo, se o bilhete 1 não será sorteado, isto é, $\neg A_1$. Como vimos, nossa loteria possui 100 mundos possíveis: $W = \{w_1; \ldots; w_{100}\}$. O mundo que nos interessa aqui é o mundo onde o bilhete 1 vence. Para facilitar, vamos dizer que isso ocorre em w_1. Desse modo, nosso agente com interesse específico particiona o universo de possibilidades de uma forma menos refinada (*fine-grained*) do que no exemplo anterior: $\{\{w_1\}; \{w_2; \ldots; w_{100}\}\}$. Ou seja, o mundo em que o bilhete 1 vence está disputando contra o conjunto dos mundos em que algum outro bilhete vence. Dado essa partição, nossa função de probabilidade terá a seguinte forma: $P(\{w_1\}) = \frac{1}{100}$; e $P(\{w2; \ldots; w_{100}\}) = \frac{99}{100}$. Nesse caso, existem dois conjuntos P-estáveis, duas opções para B_W. Uma delas é simplesmente acreditar na proposição W, ou seja, a proposição de que algum bilhete será o vencedor. A outra opção é acreditar no conjunto $\{w_2; \ldots; w_{100}\}$ de mundos, que significa acreditar na falsidade de $\{w_1\}$, isto é, acreditar que o bilhete 1 não será o vencedor. Tal contexto está fortemente enviesado contra $\{w_1\}$, e tal partição permite a crença no bilhete perdedor.

Se $TH_{\mathcal{Y}}$ estiver correta, precisamos assumir que crença racional é fortemente sensível ao contexto do agente, por meio da seleção de questões relevantes (semelhante ao que foi visto no Capítulo 2 com a teoria de Isaac Levi). Assim, uma questão deve ser entendida como uma partição de mundos onde respostas à questão correspondem a células da partição. Como vimos no final do capítulo anterior, essa parece uma consequência praticamente inevitável para qualquer princípio ponte que pretenda ser dedutivamente fechado e consistente. No exemplo anterior, evitamos o paradoxo da loteria, uma vez que não podemos colocar a crença de que o bilhete 1 não irá vencer em conjunção com outras crenças em outros bilhetes perdedores, uma vez essa crença é resultado de uma partição de um contexto específico, e não podemos aplicar regras de conjunção em crenças de contextos diferentes. Logo, temos aqui um PP que é fechado dedutivamente, é consistente, e conecta teoria da probabilidade com a lógica da crença.

3.1.3 Uma versão estável da Tese Lockeana

Outra consequência interessante é que, como mostrou Leitgeb (2015, 2017), a $TH_{\mathcal{Y}}$ é compatível com uma instância da *TL*. Vamos relembrar o que diz a Tese Lockeana:

$$Bel(A) \text{ sse } P(A) \geq s \text{ (onde } 1/2 < s \leq 1)$$

O primeiro passo para um novo entendimento sobre *TL* será notar que sua formulação é ambígua no que diz respeito aos quantificadores sobre o limiar *s* e a distribuição de probabilidade *P*. De acordo com uma possível interpretação, *TL* deve ser lida como "existe um limiar $s \leq 1$ tal que, para todo *P*,...". Por outro lado, podemos ler *TL* como afirmando "para todo *P*, existe um limiar $s \leq 1$, tal que ...". Segundo Leitgeb (2017, p. 114), essa diferença é crucial, pois enquanto a primeira interpretação é incompatível com os postulados para a crença e probabilidade subjetiva que estamos assumindo aqui, existe uma leitura da segunda interpretação que é, como veremos, compatível com fechamento lógico para crença.

O campo de escolhas permissíveis do limiar na Tese Lockeana depende agora do conjunto de crença [**B**] e de uma particular distribuição de probabilidade *P*, e não mais independentemente, como ocorre em *TL* originalmente. O núcleo de crenças B_W gerado a partir da condição de estabilidade irá determinar a abrangência de um novo limiar s^*, agora dependendo de uma particular distribuição *P*. Temos assim uma nova versão de *TL*:

TEOREMA 3.1.3 (*TL* reconsiderada)**.** *Sendo A uma proposição qualquer, Bel e P satisfazendo $TH_{\mathcal{Y}}$, então a seguinte instância de TL se sustenta: Para todo P, existe um limiar $s \leq 1$ tal que*

$$Bel(A) \text{ sse } P(A) \geq s^* = P(B_W).$$

Algumas considerações sobre o novo limiar precisam ser feitas aqui. No caso acima, é importante que o limiar seja estabelecido menor ou igual a um, uma vez que existirão casos extremos onde o único conjunto *P*-estável seja o próprio *W*, de modo que o agente só irá acreditar em proposições com probabilidade máxima. Outro fato importante é a relação do limiar s^*, ou seja, de $P(B_W)$, com os núcleos de crença: na medida em que a estrutura de núcleos de crença garante que $Bel(A)$ sse $B_W \subseteq A$, temos que qualquer proposição acreditada terá probabilidade maior ou igual a B_W, que é exatamente a propriedade de um limiar de aceitação[15].

É com esse limiar que uma instância de *TL* se segue de $TH_{\mathcal{Y}}$. Uma vez que abandonamos a ideia de que o limiar na Tese Lockeana pode ser determinado arbitrariamente – em particular, independente do modo como é *P* – então, não há nada de contraditório sobre fechamento lógico da crença, de modo que $TH_{\mathcal{Y}}$ e *TL* são satisfeitas simultaneamente. Segundo Leitgeb (2017, p. 194), o que $TH_{\mathcal{Y}}$ adiciona à *TL* é precisamente a estabilidade que é exigida à *TL* para satisfazer o fechamento lógico para crença.

Para tentar tornar esse resultado mais claro ao leitor, voltemos ao exemplo anterior, olhando para as tabelas 1 e 2. Chegamos a conclusão que, dado uma distribuição *P* sobre *W*, temos 3 opções para B_W que satisfazem $HT_{\mathcal{Y}}$: $B_W = \{w_1, w_2\}$ ou $B_W = \{w_1, w_2, w_3, w_4\}$, ou $B_W = \{w_1, \ldots, w_6\}$. Desse modo, cada núcleo será portador das seguintes probabilidades:

$$\begin{aligned} B_W = \{w_1, w_2\} &\quad = 0.72 \\ B_W = \{w_1, w_2, w_3, w_4\} &\quad = 0.97 \\ B_W = \{w_1, \ldots, w_7\} &\quad = 1 \end{aligned}$$

Assim, $TH_{\mathcal{Y}}$ acarretará três opções de limiar para nossa *TL* reconsiderada. Segundo Leitgeb (2017, p. 98), a escolha por qual limiar escolher será relativo à inclinação do agente em relação ao risco de acreditar em algo falso ou de ser demasiadamente cético. $B_W = 1$ é claramente a opção mais cética, e será a menos informativa, o que representaria um agente demasiadamente prudente. Por outro lado, $B_W = \{w_1, w_2\} = 0.72$ é a opção mais informativa, e o agente possuirá maior inclinação ao risco de erro. Nesse caso, o agente acreditará em qualquer proposição com probabilidade maior que 0.72. Obviamente, se $B_W = \{w_1, w_2\} = 0.72$, para qualquer $A \supseteq B_W$, segue-se $P(A) \geq P(B_W)$, e por consequência voltamos a nossa propriedade inicial de núcleos: $Bel(A)$ *sse* $A \supseteq B_W$.

Fica claro, desse modo, como a própria estrutura de núcleos de crença pode guiar *TL* em direção a uma versão compatível com fechamento lógico. E a relevância dos núcleos de crença não se limita a esse fato. Para Leitgeb (2017, p. 125), a estrutura dos conjuntos *P*-estáveis que originam as várias opções para B_W se assemelham aos “sistemas de esferas” de David Lewis (1973) em sua semântica para contrafactuais. Isso significa que o conjunto dos núcleos B_W é bem-ordenado no que diz respeito a relação de subconjunto, e o núcleo de probabilidade 1 é superconjunto próprio de todos os outros. Como consequência, terminamos com um ranqueamento de mundos pertencentes aos núcleos, como vimos no exemplo acima. E, como veremos no Capí-

[15] Mais detalhes sobre esse resultado e sua demonstração completa são encontradas em Leitgeb (2017, cap. 3).

tulo 4, tal estrutura possui uma implicação direta sobre o modo como entendemos a lógica da revisão de crença de modo dinâmico.

3.1.4 Uma versão estável da Teoria da Decisão Cognitiva

Outra propriedade interessante que $TH_{\mathcal{Y}}$ possui é sua compatibilidade com a Teoria da Decisão Cognitiva (*TDC*). Primeiramente, vamos retomar alguns conceitos do Capítulo 2. Se A é uma proposição acreditada, então a utilidade epistêmica de $Bel(A)$ no mundo w_i será positiva se w_i pertencer ao conjunto A, isto é, $w_i \in A$, e será negativa se w_i não pertencer ao conjunto A, isto é, $w_i \notin A$, e a utilidade epistêmica de suspender o juízo sobre uma proposição possui valor 0. Novamente, tal utilidade epistêmica positiva será representada por um número real T, onde $T > 0$, e a utilidade epistêmica negativa será representada por F, com $F < 0$.

Ainda recapitulando, o valor absoluto de $|T|$ e $|F|$ será tal que $|F| \geq |T|$, ou seja, a punição em utilidade para o erro é maior do que o benefício do acerto. Assim, a razão $\frac{-F}{T-F}$, ou de forma equivalente, $\frac{|F|}{|F|+|T|}$, deve ser interpretada como a medida de prudência do agente no que diz respeito à crença em falsidades, de modo que, quanto maior o valor de $\frac{|F|}{|F|+|T|}$, mais peso o agente atribui a preocupação sobre a falsidade de suas crenças, tornando-se mais prudente.

Como anteriormente, na medida em que não sabemos qual dos mundos possíveis é o atual, precisamos medir, de uma perspectiva puramente interna, qual a melhor atitude doxástica nessa situação de incerteza entre mundos. Novamente, para saber qual atitude adotar frente a uma proposição A, calculamos sua *utilidade epistêmica esperada*, ou sua *Acurácia Esperada*:

$$AE_{\mathfrak{a}}(b, A_i) = \sum_{w_i} P(w_i) \cdot \mathfrak{a}_{w_i}(b, v(A_i))$$

Assim, a preferência do agente por proposições verdadeiras é definida a partir da expectativa do agente em acertar, e de quanto ele valoriza a informação e desvaloriza o erro. Até aqui, nenhuma novidade em relação ao que foi apresentado anteriormente.

Avançando um pouco mais sobre o segundo capítulo, Seção 2.3.1, vimos que a razão $\frac{|F|}{|F|+|T|}$ pode ser entendida como um limiar lockeano, que irá variar conforme os valores de $|F|$ e $|T|$. Por consequência, uma Teoria da Decisão Cognitiva (*TDC*) acarreta uma instância de *TL*. Agora podemos nos perguntar: Se *TDC* é compatível com uma instância de *TL*, e $HT_{\mathcal{Y}}$ também é compatível com uma instância de *TL*, será essa uma e a mesma instância de *TL*? Não é difícil ver que isso é o caso. Leitgeb mostrou que o mesmo núcleo de crenças B_W gerado por $TH_{\mathcal{Y}}$, que gera um limiar lockeano alternativo, possui utilidade epistêmica positiva. E, como toda crença é gerada a partir de B_W, toda crença de um agente humeano possui utilidade epistêmica positiva. Portanto, daqui em diante vamos nos restringir a utilidade epistêmica da menor proposição acreditada, ou seja, o núcleo B_W.

Conforme Leitgeb (2017, p. 246), $HT_{\mathcal{Y}}$ garante que o núcleo B_W possui utilidade epistêmica positiva sempre que $r \geq \frac{|F|}{|F|+|T|}$, ou seja, para qualquer proposição A acreditada, isto é, $A \supseteq B_W$, a utilidade epistêmica de acreditar em A é maior do que a utilidade epistêmica de acreditar em sua negação, ou de suspender o juízo sobre A. Assim, se $r = 1/2$, temos que $|T| = |F|$, ou seja, $F = -T$. Como $P(\neg B_W) = 1 - P(B_W)$, podemos calcular facilmente a utilidade epistêmica do núcleo B_W. Temos que:

$$T \cdot P(B_W) - T \cdot (1 - P(B_W)) = T \cdot (2 \cdot P(B_W) - 1),$$

que é positivo sempre que $P(B_W) > 1/2$, e negativo quando $P(B_W) < 1/2$. E isso era precisamente o que vimos com Hempel (1962) ao longo do segundo capítulo: uma proposição deve ser aceita se sua probabilidade for maior que 1/2 e rejeitada se menor que 1/2. Contudo, conforme mostrou Leitgeb, podemos expandir esse resultado para outros valores de r.

Segundo Leitgeb (2017, p. 245), a suposição do parágrafo anterior de que $r \geq \frac{|F|}{|F|+|T|}$ significa que tal razão entre os valores de $|F|$ e $|T|$ funciona como um limitante inferior para o que seria um limiar admissível r na *Tese Humeana*. Em outras palavras, se a tendência psicológica do agente relativa a sua aversão ao risco de errar e sua vontade de informar está refletida no valor do limiar humeano r, isso significa que $r = 1/2$ se, e somente se, $|T| = |F|$, isto é, se o agente busca a verdade com o mesmo ímpeto com que evita a falsidade. E como vimos no capítulo 2, isso reflete um agente maximamente ousado em sua vontade de informar, assumindo que ele deseja preservar consistência. Isso significa que o limiar $r = 1/2$ que viemos assumindo até aqui para $TH_{\mathcal{Y}}$ é bastante limitado em refletir a diversidade de tendências psicológicas dos agentes epistêmicos. E como veremos no final deste capítulo, Seção 3.3.2, assumir que $r = 1/2$ implica em outras vantagens e desvantagens para $TH_{\mathcal{Y}}$.

Agora é importante avançar um pouco mais em $TH_{\mathcal{Y}}$ para explorar resultados com valores de r maiores que 1/2. Para isso, precisamos de um adendo na formulação inicial da condição de preponderância (como já mencionado na nota 13). Assim, se o núcleo B_W satisfaz a condição de preponderância, então $P(B_W) > \frac{r}{1-r} \cdot (1 - P(B_W))$. Calculando os valores de utilidade epistêmica esperada de $Bel(B_W)$ a partir dessa desigualdade, temos o seguinte:

$$T \cdot P(B_W) + F(1 - P(B_W)) > T\frac{r}{1-r}(1 - P(B_W)) + F(1 - P(B_W)) = (T\frac{r}{1-r} + F)(1 - P(B_W)) \tag{12}$$

Conforme vimos acima, uma exigência para o núcleo B_W possuir utilidade epistêmica positiva é que $r \geq \frac{|F|}{|F|+|T|}$, que é equivalente a $\frac{1-r}{r}|F| \leq |T| = T$. Com isso em mãos, da última igualdade em 12, substituindo T por $\frac{1-r}{r} \cdot |F|$ temos que

$$(T\frac{r}{1-r} + |F|)(1 - P(B_W)) \geq (\frac{1-r}{r} \cdot |F| \cdot \frac{r}{1-r} + F)(1 - P(B_W)) \tag{13}$$

Como $(\frac{1-r}{r} \cdot \frac{r}{1-r}) = 1$ e, dado $F < 0, |F| + F = 0$, temos que

$$(\frac{1-r}{r} \cdot |F| \cdot \frac{r}{1-r} + F)(1 - P(B_W)) = 0.$$

Portanto, para qualquer valor r do limiar humeano, B_W possuirá utilidade epistêmica estritamente maior que 0, e consequentemente, qualquer proposição acreditada por $TH_{\mathcal{Y}}$ possui utilidade epistêmica maior do que a utilidade de suspender o juízo sobre ela, uma vez que, como vimos, $Bel(A)$ *sse* $A \supseteq B_W$.

Por fim, é importante ressaltar que uma das vantagem de $TH_{\mathcal{Y}}$ em relação à *TDC* é seu caráter simétrico: dada uma distribuição de probabilidade subjetiva e um conjunto de crenças, posso resolver um caso de inconsistência entre elas revisando qualquer uma das parte. Já com a *TDC*, como vimos no final do segundo capítulo, existe claramente uma prioridade da racionalidade da probabilidade subjetiva sobre a racionalidade da crença. E como vimos no primeiro capítulo, estamos buscando por um princípio ponte que seja um requisito racional estrutural, e a simetria entre estados é uma condição necessária para tais requisitos.

3.2 A TEORIA PROBALÓGICA

Antes de entrar precisamente no assunto deste tópico, vamos retomar alguns dos pontos que vimos até então. Dado os resultados do Capítulo 2, estabelecemos que a rota de investigação na busca por um princípio ponte que conecte crença (*simpliciter*) e probabilidade subjetiva seria assumindo o fechamento lógico-dedutivo como um requisito da racionalidade para crença, mesmo que isso custasse a aceitação de uma concepção de crença altamente dependente do contexto. Assim, no presente capítulo foi estabelecido que continuaríamos nossa busca por um PP a partir de duas teses desenvolvidas recentemente que satisfazem os postulados da crença e da probabilidade subjetiva, conforme vimos no primeiro capítulo. A *Tese Humeana* da crença foi a primeira delas. Enquanto na *Tese Lockeana*, para determinar se uma determinada proposição A deve ou não ser acreditada precisamos considerar a probabilidade absoluta de A, em *TH* o que importa é a probabilidade condicional de A, isto é, a probabilidade de A dado algum outro evento. Agora veremos como um PP pode ser estabelecido levando em consideração, não a probabilidade absoluta ou condicional de uma proposição, mas sua razão/cotação entre proposições. E isso ocorrerá por meio da *Teoria Probalógica*, desenvolvida pela dupla Hanti Lin e Kevin Kelly (2012; 2012).

Relativamente à *TH*, a *Teoria Probalógica* (*PL*) é mais limitada no que diz respeito à amplitude de seu desenvolvimento teórico. Por isso sua apresentação aqui será mais breve comparada a sua rival humeana. Todavia, os trabalhos de Lin e Kelly foram fundamentais para a construção de uma abordagem formal para regras que cumpram o papel de um PP, principalmente no que diz respeito a uma interpretação geométrica de tais regras (como veremos no próximo, e último, tópico deste capítulo). E ainda, como veremos no próximo capítulo, a *PL* possui resultados extremamente importantes no que diz respeito a construção de um PP dinâmico, estabelecendo uma disputa teórica com a *Tese Humeana* sobre qual a melhor teoria para revisão de crença que satisfaz as condições relevantes. Mas isso será assunto para o próximo capítulo.

Vamos iniciar estabelecendo alguns dos pressupostos teóricos subjacentes à teoria. Como de costume, partimos de um conjunto W de mundos possíveis e uma álgebra de conjuntos $\mathcal{A}$ formada a partir do conjunto das partes $\mathscr{P}(W)$. Dizemos que uma *questão relevante* $\mathcal{Q}$ (semelhante ao que vimos no capítulo 2 com Levi) particiona o conjunto $\mathcal{A}$ de proposições em um conjunto H_i de possibilidades mutuamente excludentes e coletivamente exaustivas. Cada célula $H_i \in \mathcal{Q}$ representará uma, e somente uma, resposta completa para $\mathcal{Q}$. Para facilitar nosso trabalho, daqui em diante vamos assumir que cada resposta completa H_i é um conjunto unitário de mundos possíveis, ou seja, para cada $H_i \in \mathcal{Q}$ e cada $w_i \in W$, temos que $H_i = \{w_i\}$, e qualquer união $\{w_i\} \cup \{w_j\} \cup ... \cup \{w_n\}$ é uma resposta parcial, e menos informativa, à questão relevante de tal modo que, quanto maior for o número de elementos disjuntos na resposta, mais cética e menos informativa ela será. Por fim, definimos $\mathcal{P}$ como o conjunto de todas distribuições de probabilidade definidas sobre W, e por consequência, o conjunto de todas distribuições de probabilidade possíveis sobre as respostas em $\mathcal{Q}$.

Ainda sobre o conjunto de pressupostos, como de costume, dada uma distribuição de probabilidade P sobre W, atribuímos para cada w_i uma medida de probabilidade $P(w_i)$. Agora precisamos introduzir algumas novas noções que nos serão úteis daqui em diante. Primeiramente, dada uma distribuição de probabilidade P, dizemos que P acarreta uma determinada proposição A_i quando o PP garante que a $P(A_i)$ é suficiente para tornar A_i uma proposição aceita no conjunto de crenças (seja A_i apenas um conjunto unitário — uma resposta completa — ou uma união de conjuntos unitários — uma resposta parcial). Ainda, seja Bel_P o conjunto de todas as crenças acarretadas por uma determinada distribuição P. Dizemos que $Bel_P(A_i)$ se, e somente se, P acarreta A_i de acordo com um PP específico. Por fim, dizemos que uma regra de aceitação é *consistente* se não existe uma distribuição P tal que $Bel_P(\emptyset)$, e dizemos que é *não-cética* se, para cada resposta parcial $\{w_i\} \cup ... \cup \{w_n\}$, existe uma distribuição de probabilidade P tal que $P(\{w_i\} \cup ... \cup \{w_n\}) < 1$ e $Bel_P(\{w_i\} \cup ... \cup \{w_n\})$.

Como mencionado anteriormente, a principal estratégia dos autores é tomar não mais o valor de probabilidade absoluto de cada proposição em relação a um limiar, mas agora tomar a probabilidade de cada resposta completa em uma relação de proporção com a probabilidade de uma outra resposta. Uma das principais diferenças de *PL* é que, agora, crença não será mais determinada aos moldes de um limiar ao estilo lockeano. Para Lin e Kelly (2012b, p. 966), o limiar para um PP adequado deve ser baseado no que eles chamam de “limiar de cotação” (do inglês *odds threshold*).

Em teoria da probabilidade, uma “cotação” (*odds*) é uma razão entre a chance de um evento ocorrer ou não. Isso é geralmente utilizado em contextos de apostas onde queremos saber o quão mais provável é o evento A em relação ao evento $\neg A$, de modo que uma cotação na forma 1:3 contra A significa que $P(A) = 0.75$. Em outras palavras, um agente racional estaria disposto a apostar $25R\$$ contra o evento A, ou seja, não é o caso que A, para ganhar $100R\$$.

Resumidamente, um limiar de cotação deve ser construído levando em conta a probabilidade de uma resposta completa à questão relevante em relação a uma resposta concorrente. Mais precisamente, o que deve ser levado em conta é a probabilidade de uma resposta em relação a sua concorrente mais provável. Para Lin e Kelly (2012b, p. 967), um limiar de cotação será gerado a partir de uma *ordem de plausibilidade* entre mundos: para cada $w_i \neq w_j \in W$, temos que w_i é estritamente mais plausível[16] que w_j se, e somente se, $\frac{P(w_i)}{P(w_j)} > t$, para algum valor de t estritamente maior que 1. Nesse sentido, se estamos interessados na resposta representada pela proposição $\{w_i\}$, devemos comparar a probabilidade de w_i com a resposta concorrente mais provável, ou seja, $\frac{P(w_i)}{\max_j P(w_j)}$[17]. Tal razão de proporção originará um limiar de cotação para rejeição de respostas em PL, como veremos a seguir.

Basicamente, indo diretamente ao ponto, PL determina um conjunto de crenças a partir da seguinte regra (LIN; KELLY, 2012a, p. 537)[18]:

DEFINIÇÃO 23 (Teoria Probalógica)**.** Seja $t > 1$, e $w_i, w_j \in W$ respostas completas para $\mathcal{Q}$:

$$Bel(\neg w_i) \ sse \ \frac{P(w_i)}{\max_j P(w_j)} \leq \frac{1}{t}$$

Uma forma equivalente de formular tal regra é tomar o conjunto de crenças do agente como sendo o complemento da intersecção de todas respostas rejeitadas. Formalmente,

$$Bel(A) \ sse \ A \supseteq \left\{ W \setminus w_i : \frac{P(w_i)}{\max_j P(w_j)} \leq \frac{1}{t} \right\}$$

Como veremos no tópico 3.2, esse princípio de aceitação é conhecido como

16 No próximo capítulo iremos explorar mais sobre tal relação de ordem e sua relevância para uma teoria da revisão de crença.

17 Para simplificar a notação, daqui em diante falaremos apenas de mundos como candidatos a respostas completas para $\mathcal{Q}$, e os colchetes em $\{w_i\}$ serão omitidos. Mas é importante enfatizar que quando falarmos de mundo como respostas, estamos nos referindo à proposição relevante, a resposta completa, que tal mundo torna verdadeira. O mesmo ocorre quando falarmos de “disjunção” e “conjunção” de mundos possíveis.

18 Para os nossos propósitos, a formulação aqui é ligeiramente distinta da original apresentada pelos autores, tanto em Lin e Kelly (2012b, p. 970) quanto em Lin e Kelly (2012a, p. 537). Ainda, o modo formalmente adequado de apresentar a definição abaixo seria utilizar $Bel(\neg\{w_2\})$ como a crença no complemento da proposição unitária que contém w_2. Como já dissemos, os colchetes são omitidos pela simplicidade da apresentação.

"regra do obturador", devido a sua forma geométrica[19]. A origem dessa regra se encontra nos trabalhos de Isaac Levi (2007, p. 286), embora ele mesmo rejeite a regra para seus propósitos específicos. No restante deste capítulo, e no próximo, pretendo deixar clara a intuição por trás desse princípio e quais elementos novos ele traz para o nosso problema. Por enquanto, vamos ver como ele se aplica aos nossos desafios anteriores.

Retomemos nosso exemplo anterior, a partir da Tabela 1 com oito mundos possíveis. Nesse caso, teremos 8 mundos descrevendo cada uma das possibilidades. Como dois dos mundos possuem probabilidade zero, eles serão ignorados. Essa será nossa questão relevante $\mathcal{Q}$, onde cada w_i representa uma resposta para $\mathcal{Q}$. Queremos agora saber o que o PP Probalógico nos diz sobre qual resposta deve ou não ser aceita. Pela Definição 23, tal regra determinará qual mundo deve ser rejeitado como resposta. Aplicando a regra sobre os mundos em W, conforme a Tabela 2, por fechamento lógico, teremos a conjunção de todas as respostas rejeitadas, e poderemos derivar desse resultado qual a resposta mais forte que deve ser aceita.

Em nosso exemplo, a Tabela 2 mostra que $\max_j P(w_j) = P(w_1) = 0.4$. Assim, se queremos saber se w_i estará entre o conjunto de respostas para $\mathcal{Q}$, para qualquer $w_i \in W$, basta aplicar a regra Probalógica sobre $P(w_i)$ comparado à $\max_j P(w_j) = P(w_1)$. Como nos princípios anteriores, o valor para o limiar t é uma constante que deve ser ajustada conforme a questão relevante. Por exemplo, dado um limiar $t = 2$ (o que significa uma cotação de 1:2 entre as probabilidades comparadas), se queremos saber se w_2 é uma resposta para $\mathcal{Q}$ segundo a regra probalógica, devemos primeiramente testar se seu complemento é aceito. Assim, temos que

$$Bel(\neg w_2) \ sse \ \frac{0.32}{0.4} \leq \frac{1}{2},$$

ou seja, a regra do obturador diz que o complemento de w_2 *não deve* ser aceito como resposta à questão relevante, uma vez que $\frac{0.32}{0.4} = \frac{4}{5} > \frac{1}{2}$. Assim, com um limiar $t = 2$, w_2 não foi excluído do conjunto de crenças. Se queremos saber se w_3 será rejeitada sob o mesmo limiar, basta tomar

$$Bel(\neg w_3) \ sse \ \frac{0.14}{0.4} \leq \frac{1}{2},$$

ou seja, um PP Probalógico diz que o complemento de w_3 deve ser aceito como resposta à questão relevante, uma vez que $\frac{0.14}{0.4} = \frac{7}{20} < \frac{1}{2}$. Assim, w_3 não será parte do conjunto de respostas aceita pelo agente. E como $P(w_3) \geq P(w_4) \geq \cdots \geq P(w_6)$, é fácil perceber que todos os mundos após o terceiro serão rejeitados, o que acarreta:

$$Bel(\neg w_3 \wedge \neg w_4 \wedge \cdots \wedge \neg w_6).$$

[19] "Obturador" é o nome que se dá ao mecanismo existente em câmeras fotográficas que se abre e fecha protegendo o sensor da câmera, abrindo somente no momento em que o disparador é acionado, permitindo assim a entrada de luz (também conhecido como " olho da câmera").

Logo, em nosso exemplo, $Bel(w_1 \vee w_2)$ será a única resposta aceita, e o conjunto de crenças do agente será o mesmo conjunto em $TH_{\mathcal{Y}}$: o agente acredita na proposição de que a pizzaria abre aos sábados e não fecha antes das 23:30.

Agora suponha que nosso agente seja mais prudente, de modo que $t = 5$. Utilizando novamente os valores do nosso exemplo na Tabela 2, nesse caso, w_3 não será rejeitada, uma vez que $\frac{0.14}{0.4} = \frac{7}{20} > \frac{1}{5}$. O mesmo ocorre com w_4, dado que $\frac{0.11}{0.4} \approx \frac{3}{10} > \frac{1}{5}$. Contudo, isso já não ocorre com o mundo w_5, pois $\frac{0.02}{0.4} = \frac{1}{20} < \frac{1}{5}$. Nesse caso, o conjunto de crenças do agente seria a disjunção dos 4 mundos mais prováveis. No exemplo, a única resposta acreditada pelo agente seria a de que a pizzaria abre aos sábados, permanecendo cético sobre seu horário de fechamento.

Como isso nos ajuda? Bom, é fácil ver como isso soluciona o paradoxo da loteria. Assim como em $TH_{\mathcal{Y}}$, a teoria Probalógica também é dependente do modo como a questão relevante $\mathcal{Q}$ é selecionada por meio de uma partição. Isso faz com que num contexto de uma loteria justa com mais de três bilhetes, nenhuma proposição sobre a derrota de um bilhete específico se destaque entre as respostas concorrentes, tal que $\frac{P(w_i)}{\max_j P(w_j)} \leq \frac{1}{t}$, dado $t > 1$. Mas isso poderia ser diferente se a questão relevante $\mathcal{Q}$ for particionada w_1 e $w_2, \ldots w_n$. Nesse caso, a partição está fortemente enviesada para o complemento de w_1, uma vez que $P(w_1) = \frac{1}{n}$ e $P(\neg w_1) = \frac{n-1}{n}$. Na Seção 3.2 veremos como essa regra oferece outros elementos interessantes para lidar com o paradoxo da loteria quando olharmos para o paradoxo de um ponto de vista geométrico.

É importante notar uma diferença fundamental entre *PL* e suas rivais $TH_{\mathcal{Y}}$ e *TL*: uma vez que *PL* não funciona na forma de limiares de aceitação ela possui algumas consequências que, ao menos em um primeiro momento, soam um tanto quanto contraintuitivas. Primeiramente, para uma distribuição de probabilidade uniforme (isto é, $P(w_i) = P(w_j)$ para todo $w_i, w_j \in W$), *PL* implica que nenhuma resposta não-trivial será aceita, por maior que seja sua probabilidade, e independentemente do valor do limiar t. Em segundo lugar, *PL* permite que uma resposta seja aceita por menor que seja sua probabilidade, e independentemente do valor de t. Para ver como isso ocorre, basta assumir um W com um grande número de mundos de tal modo que $P(w_i)$ seja extremamente pequena, mas ainda assim suficientemente maior que suas rivais, o suficiente para garantir que o limiar de cotação rejeite todas suas concorrentes. Uma consequência disso é que, por exemplo, devemos acreditar que fulano será o vencedor em uma loteria com um milhão de bilhetes simplesmente pelo fato de ele ser a única pessoa que comprou dois bilhetes.

Uma consequência do que vimos no parágrafo acima é que, apesar de *PL* fornecer uma resposta ao paradoxo da loteria, a teoria é incapaz de oferecer uma resposta ao paradoxo do prefácio (Seção 1.3.4), uma vez que alta probabilidade não é necessária para crença. Todavia, em seu trabalho mais recente (LIN; KELLY, 2021, p. 199), a dupla Lin e Kelly adiciona uma exigência de alta probabilidade entre os requisitos de sua teoria: Se é o caso que $Bel(A)$ em PL, então $P(A) > 1/2$. Para isso, a seguinte condição deve ser adicionada sobre o valor do limiar t que aparece na definição 23:

$$\sum_{w_i} \frac{1}{t_{w_i}} < 1$$

Tal exigência sobre o valor de t garantirá que, no caso do parágrafo anterior em que W possui um grande número de mundos, e consequentemente $\mathcal{Q}$ um grande número de respostas mutuamente excludentes, o valor de t seja grande o suficiente para fazer com que a soma de $\frac{1}{t_{w_i}}$, tantas vezes quanto for o número de mundos $w_i \in W$, seja estritamente menor que 1. Por exemplo, quando $|W| = 3$, temos que o valor de t deve ser tal que $\frac{1}{t} + \frac{1}{t} + \frac{1}{t} < 1$, ou seja, t deve ser estritamente maior do que 3. Com isso, PL evita parte das objeções do parágrafo anterior, garantindo alta probabilidade seja necessária para crença.

Por fim, podemos afirmar que o grande mérito da teoria Probalógica até aqui foi ter oferecido uma regra de aceitação para crença que cumpre fechamento lógico, satisfaz os postulados para probabilidade subjetiva, e ainda assim oferece um conjunto de crenças não-cético e consistente. No próximo tópico veremos como não-ceticismo e consistência para um conjunto de crenças pode ser entendido não mais olhando para a definição formal da regra de aceitação, mas sim a partir da forma geométrica gerada por tal regra. E é a partir dessa nova perspectiva que iremos comparar os resultados obtidos por TL, $TH_{\mathcal{Y}}$ e PL no que diz respeito as respectivas capacidades de oferecerem conjuntos de crença não-céticos e consistentes.

3.3 A GEOMETRIA DO PRINCÍPIO PONTE

Neste tópico, veremos como podemos conectar a lógica da crença e a teoria da probabilidade utilizando de princípios geométricos, e como tal abordagem pode impactar na compreensão de nosso problema relevante. Apesar dessa ideia já ter sido significativamente explorada na literatura, inicialmente com Bruno de Finetti (1965) e mais recentemente com Lin e Kelly (2012a), a apresentação a seguir oferece uma introdução mais acessível ao leitor não familiarizado com a abordagem geométrica do problema. E uma apresentação mais introdutória da questão, diga-se de passagem, é algo bastante raro na literatura relevante.

Como ponto de partida, é importante destacar que a noção de "distância" terá um destaque especial, como teve também no capítulo 2, Seção 2.2, quando falamos em teorias da acurácia. Veremos como a distância entre a verdade de uma proposição e sua probabilidade subjetiva pode ser medida por meio da distância entre os vértices de um triângulo equilátero. Mas vamos com calma. Antes de entrar especificamente em nosso objeto alvo, ou seja, nos modelos geométricos de PP, vamos apresentar

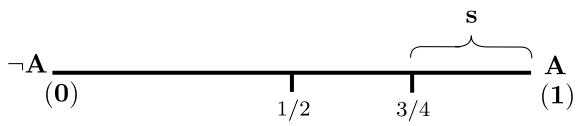

Figura 1 – Segmento de reta unitário contendo a proposição A e sua negação.

alguns dos *insights* iniciais por trás da relação geométrica que podemos estabelecer entre probabilidade e lógica.

Como vimos no Capítulo 2, podemos avaliar a relação entre probabilidade e verdade a partir de uma medida de distância em um segmento de reta. Na Figura 1, podemos ver um segmento de reta, de comprimento unitário, com cada um dos extremos representando, respectivamente, a verdade e a falsidade de uma determinada proposição $A \in \mathcal{A}$. Assim, uma medida de probabilidade será um ponto na reta unitária, ou seja, um elemento do intervalo $[0, 1]$. Um limiar $s = 3/4$ na reta determinará quando uma distribuição de probabilidade acarretará, em *TL*, a crença em uma determinada proposição. Em outras palavras, dado um limiar s em um segmento de reta unitário, para qualquer ponto dentro do intervalo $[0, 1]$ representando a probabilidade subjetiva do agente na proposição, *TL* diz se essa proposição será ou não aceita. Desse modo, somos capazes de representar alguns dos conceitos mais relevantes em nosso trabalho, como "proposição", "verdade/falsidade", "probabilidade", e "limiar de aceitação", a partir de noções matemáticas/geométricas bem elementares, como segmento de reta, ponto, intervalo e distância. Isso pode ser notado a partir da Figura 1 (no caso de proposições mutuamente excludentes, $\neg A = B$).

Conforme a Lógica da crença, temos que o número de combinações de valores de verdade e falsidade para um conjunto de proposições será de 2^n, sendo n o número de proposições em questão. Logo, com apenas uma proposição, temos 2^1 combinações, ou seja, A e $\neg A$. Para ampliar nossa compreensão, vamos a mais alguns casos. Quando temos duas proposições, teremos 2^2 combinações. Assim, se quisermos visualizar isso geometricamente, temos que passar do $\mathbb{R}^1$ para o $\mathbb{R}^2$. A Figura 2 mostra a expansão do segmento de reta anterior para o plano, agora com duas proposições A e B. Na figura em questão, uma distribuição de probabilidades será um ponto no plano; a região cinza claro representa a área de aceitação da Tese Lockeana com limiar $s = 0.75$ para a proposição mais forte $A \wedge B$.

O leitor familiarizado com teoria da probabilidade já deve ter notado como tal formulação geométrica reflete perfeitamente os axiomas probabilísticos: a intersecção de proposições sempre reduz a região para a aceitação da conjunção $A \wedge B$, enquanto a união de proposições aumenta a região da conjunção de $A \vee B$. Assim, quanto maior o número de proposições, menor será a área geométrica ocupada pela conjunção de todas as proposições, ou seja, menor será a probabilidade da intersecção.

E no caso de 3 proposições? Nesse caso, teremos 2^3 combinações lógicas pos-

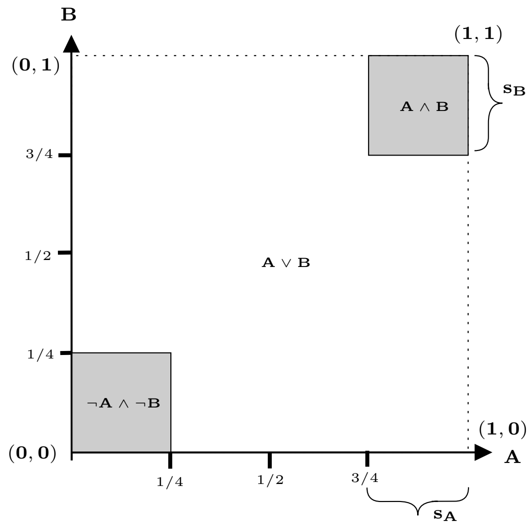

Figura 2 – Plano representando as combinações lógicas das proposições *A*, *B*.

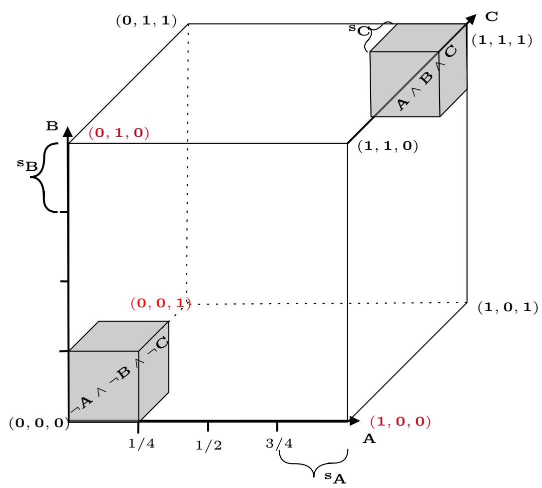

Figura 3 – 2^3 combinações lógicas das proposições *A*, *B*, *C*.

síveis, e precisaremos expandir nosso $\mathbb{R}^2$ para o $\mathbb{R}^3$. A representação de um cubo em uma tela será sempre limitada, mas ainda é possível, como podemos ver na Figura 3. Na imagem, uma distribuição de probabilidade é um ponto em alguma região no cubo. A região de aceitação com limiar $s = 0.75$ é um pequeno cubo na extremidade $(1, 1, 1)$, e será medido a partir do volume ocupado pelo pequeno cubo cinza. Novamente, a região de aceitação para a proposição mais forte $A \wedge B \wedge C$ torna-se cada vez menor conforme adicionamos proposições independentes.

Para representar geometricamente a combinação de 4 proposições, precisaríamos expandir $\mathbb{R}^3$ para $\mathbb{R}^4$, o que dificulta bastante a visualização. Contudo, formalmente, podemos expandir as dimensões para representar qualquer conjunto com n proposições utilizando formalização algébrica. Mas, felizmente, isso não será necessário aqui, uma vez que vamos trabalhar com exemplos de eventos em partições. Ao invés de olhar para todo o cubo da figura anterior, nosso foco será as coordenadas

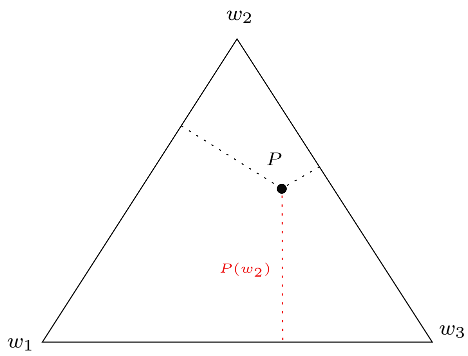

Figura 4 – Como medir a probabilidade de um evento dentro do 2 – *simplex*.

"normalizadas", isto é, as combinações cujos valores das coordenadas somam 1. Isso significa que estaremos interessados em partições de proposições mutuamente excludentes e coletivamente exaustivas. Os valores de verdade dessa partição estão destacadas de vermelho na Figura 3, em três dos vértices do cubo, especificamente, representando as três proposições em questão. Contudo, existe uma maneira mais simples e intuitiva de representar geometricamente uma medida de probabilidade em uma partição, o que é feito por meio de um 2 – *simplex* unitário[20]. E é esse caminho que iremos percorrer aqui.

Formalmente, uma distribuição de probabilidade na partição relevante representa uma medida P, para cada $w_i \in W$, com P sendo um ponto arbitrário (os infinitos pontos no simplex correspondem as infinitas medidas de probabilidade sobre a partição). Na Figura 4 podemos ver o caso em que $W = \{w_1, w_2, w_3\}$, ou se preferir, com 3 proposições mutuamente excludentes onde cada mundo pertence a uma, e apenas uma, proposição. Dada uma distribuição de probabilidade normalizada em W, P é uma distribuição tal que $P(w_1) + P(w_2) + P(w_3) = 1$. Como podemos observar na figura, a probabilidade de uma proposição será a distância entre o ponto P e uma reta perpendicular (passando pelo ponto P) à aresta oposta ao vértice que representa a proposição em questão. No exemplo da Figura 4, as linhas pontilhadas representam as respectivas medidas de probabilidade para cada proposição da partição, e a medida da linha pontilhada em vermelho será exatamente a probabilidade que P atribui a w_2[21].

Uma vantagem de representar uma distribuição de probabilidade sobre W por meio de um simplex unitário (ou simplex de probabilidade) é o fato de que a soma

[20] Um $n-$*simplex* é um subespaço de $\mathbb{R}^n$ constituindo as formas geométricas mais elementares. Um $n-$*simplex* possuirá $n+1$ vértices, de modo que o elemento geométrico mais simples, isto é, o ponto, será um 0 – *simplex*. O segmento de reta é um 1 – *simplex*; o triângulo equilátero é um 2 – *simplex*; o tetraedro é um 3 – *simplex*; e por aí vai. Os *simplex* serão extremamente úteis para representarmos distribuições de probabilidades em uma partição de $n + 1$ células.

[21] Nesse caso, a altura do triângulo é tomada como uma unidade. No caso típico onde as arestas possuem medida 1, pelo teorema de Pitágoras, sabemos que a altura é $\sqrt{1^2 - (\frac{1}{2})^2} = \frac{\sqrt{3}}{2}$. Nesse caso, a probabilidade de uma proposição é a distância entre a aresta e o ponto P, multiplicado por $\frac{\sqrt{3}}{2}$.

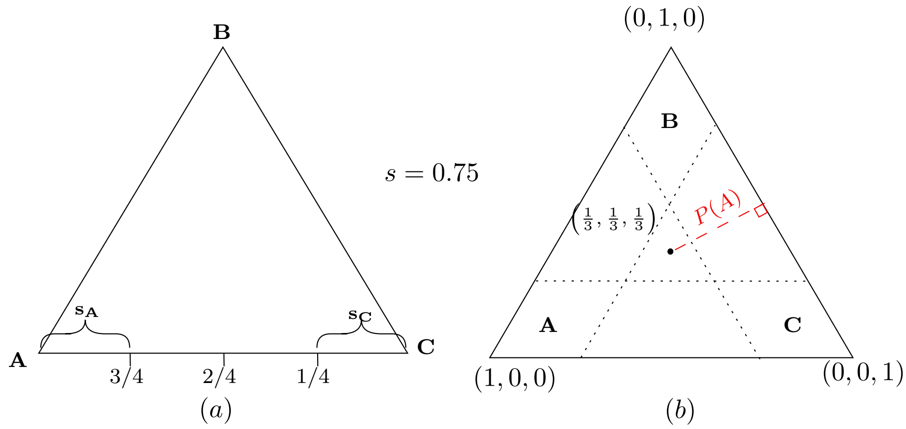

Figura 5 – Como interpretar limiares de aceitação no 2 – *simplex*.

das medidas até P (representadas pelas linhas pontilhadas na Figura 4) satisfazem a "condição de normalização". Quando W possui n mundos possíveis, e $P(w_i)$ é a distância $|w_i|$ entre o ponto P no plano e à aresta oposta ao vértice w_i, temos que:

$$\sum_{i=1}^{n} |w_i| = 1$$

Agora é importante relembrar algumas propriedades de uma distribuição de probabilidade. Se A e C são proposições[22] mutuamente excludentes, quanto maior for a probabilidade de uma, menor será a probabilidade da outra (mantendo o resto constante), ou seja, quanto mais próximo de A, mais distante de C, ou melhor, mais próximo de $\neg C$ (como ilustramos anteriormente na Figura 1). Partindo para nosso caso com três proposições em um 2 – *simplex*, quando a probabilidade de B é zero, a distribuição de probabilidade será igual a distribuição no segmento $\overline{AC}$.

Caminhando em direção ao que realmente nos interessa, vamos lembrar que o conjunto de crenças do agente é determinado na Tese Lockeana pela seguinte regra: $Bel(A)$ *sse* $P(A) \geq s$. Isso implica que $Bel(\neg A)$ *sse* $P(A) \leq 1 - s$. Com isso em mente, podemos traçar no 2–*simplex* as regiões de aceitação para cada limiar s. Podemos ver um exemplo na Figura 5 de um limiar $s = 0.75$ para as três proposições alvo, de modo que cada ponto no plano represente uma distribuição de probabilidade normalizada, e um limiar de aceitação dirá o que será ou não acreditado para cada distribuição de probabilidade $P \in \mathcal{P}$. Na Figura 5(a), a base do triângulo funciona como o intervalo unitário entre A e C. O mesmo ocorre com cada aresta do triângulo. Do lado direito, em (b), a linha pontilhada representa o limiar lockeano s de rejeição para cada proposição. O ponto no centro do triângulo é a distribuição $P(A) = P(B) = P(C) = 1/3$, e a linha pontilhada em vermelho é a probabilidade de A.

[22] Para facilitar, daqui em diante iremos nos referir a proposições (conjuntos de mundos possíveis) como as representantes de cada mundo em uma distribuição de probabilidade na partição relevante.

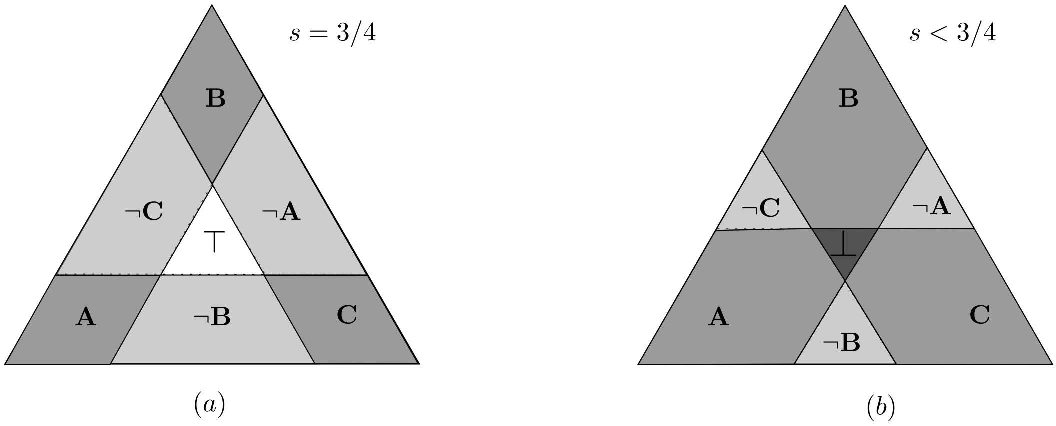

Figura 6 – Como a Tese Lockeana dá origem a uma zona onde uma proposição contraditória é aceita.

A origem dessa interpretação geométrica para medidas de probabilidade e as possíveis regiões de aceitação para as proposições relevantes pode ser encontrada principalmente em de Finetti (1965). Todavia, foram os proponentes da *Teoria Probalógica*, Lin e Kelly (2012b), que desenvolveram com maior profundidade esse tópico, e boa parte do que será apresentado a seguir é devido ao trabalho deles. Lin e Kelly (2012a) perceberam que o paradoxo da loteria possui uma forma geométrica, e o paradoxo ocorre fundamentalmente devido à geometria de *TL*. Como podemos observar na Figura 5, o PP baseado no limiar lockeano *s* pode ser interpretado geometricamente como uma região, ou zona, do plano. Isso fica mais claro na Figura 6(*a*).

Nesse caso, em uma partição com três elementos, um limiar $s = 3/4$ é suficiente para garantir que o princípio lockeano seja consistente, ou seja, nenhuma região dentro do triângulo, para qualquer distribuição *P*, acarreta um conjunto contraditório de proposições. Em outras palavras, não existe uma região do triângulo onde o ponto de distribuição de probabilidades acarretará uma contradição. Contudo, conforme o limiar $s = 3/4$ decresce em direção a $2/4$, a proposição contraditória se torna uma opção para uma determinada distribuição *P*, como podemos ver na Figura 6(*b*).

Como o leitor pode observar a partir da Figura 6 (*a*), as regiões cinza escuro representam a área de aceitação da respectiva proposição mais forte; as regiões cinza claro representam a área de aceitação para a disjunção das duas proposições, por exemplo, $A \vee B = \neg C$; e a região central representa a área onde a distribuição de probabilidade *P* irá garantir somente a aceitação da proposição certa $A \vee B \vee C$. Em (*b*), podemos ver o que acontece quando o limiar decresce. Assim, as áreas com formato de losango localizadas em cada um dos cantos representam as áreas de crença para cada resposta potencial. Conforme *s* decresce, passa a existir uma área onde as proposições *A*, *B*, *C* são todas aceitas. Mas esse conjunto de proposições é inconsistente, de modo que, por fechamento lógico, essa região de aceitação corresponde à proposição contraditória $\perp$. Em uma loteria com três bilhetes, interpretando a proposição *A* como sendo a proposição "o bilhete *A* é o bilhete vencedor", esse é basicamente o paradoxo da loteria com limiar ¾. Assim, é fácil ver como *TL* é ou cética (isto é, $s = 1$) ou inconsistente (isto é, permite que uma certa distribuição de probabilidade implique uma proposição contraditória).

Mas por que tudo isso nos interessa? Não seria apenas uma complicação desnecessária? Primeiramente, o aspecto geométrico da regra lockeana nos ajuda a buscar por novas soluções aos paradoxos. Por meio da visualização e interpretação geométrica, não é difícil pensar em uma maneira intuitiva de reformular *TL* para que o aumento do valor do limiar não implique em uma sobreposição de proposições mutuamente excludentes. Veremos como isso ocorre no restante do presente capítulo.

3.3.1 A geometria da Tese ProbaLógica

Dado o reconhecido mérito do trabalho de Lin e Kelly (2012a) neste tópico, começaremos agora a apresentar as novas interpretações geométricas de um PP, começando pela *Teoria Probalógica*. Retomemos, assim, o que nos diz a definição 23 de *PL*. Seja $t > 1$, e $w_i, w_j \in W$ respostas para $\mathcal{Q}$:

$$Bel(\neg w_i) \ sse \ \frac{P(w_i)}{\max_j P(w_j)} \leq \frac{1}{t}$$

Como já foi dito anteriormente, enquanto *TL* determina o conjunto de crenças do agente utilizando-se da probabilidade absoluta das proposições, *PL* considera as "cotações" (*odds*) entre a probabilidade de certas proposições. Na Figura 7 podemos ver o resultado de aplicar a regra anterior em nosso $2-simplex$. Em (*a*) podemos ver o $2-simplex$ para um limiar $t = 1$. Em (*b*) vemos um limiar $t = 3$. A linha vermelha em destaque é chamada de "linha de cotações constantes", e a região branca $\top$ é o espaço de aceitação cético (agora deve ficar mais claro para o leitor a razão pela qual a regra é chamada de "regra do obturador", ou "regra do olho da câmera"). Na medida em que o valor de *t* decresce, e a teoria fica cada vez menos cética, a figura (*b*) caminha em direção à figura (*a*), de modo que não ocorra um colapso entre regiões de aceitação, tal como ocorre com *TL*.

Embora a teoria esteja definida apenas para $t > 1$, podemos arbitrariamente fazer *t* tão próximo de 1 quanto desejarmos, e tal caso está representado na Figura 7(*a*). Nesse caso, a única distribuição cética é o centro do triângulo, e, diferentemente de *TL*, não existe espaço onde certa distribuição de probabilidade aceita uma proposição contraditória. Na Figura 7(*b*), vemos a forma geométrica quando $t = 3$. Nesse caso, a teoria se torna mais cética, aumentando assim o espaço ocupado pela região tautológica $\top$ (a parte branca da figura).

O principal elemento responsável pela forma geométrica de *PL* é a "linha de cotações constantes", como, por exemplo, a linha vermelha destacada em Figura 7(*b*).

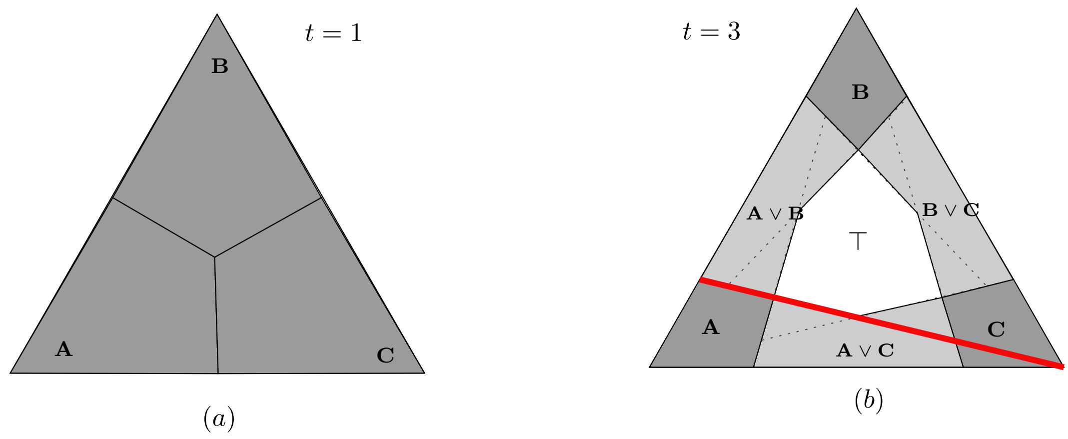

Figura 7 – A interpretação geométrica da Teoria ProbaLógica.

Nesse exemplo, a linha particiona o triângulo em duas regiões: abaixo da linha está o espaço onde a proposição *A* é preferível à proposição *B*, ambas tomadas relativamente a *C*. Em outras palavras, abaixo da linha está o espaço onde $\neg B$ é aceito relativamente a *C*. Seu nome é devido ao fato de que sobre tal linha se encontram todas as distribuições de probabilidade que garantem uma cotação de 3:1 para $\frac{P(A)}{P(C)}$ contra $\frac{P(B)}{P(C)}$. Ou seja, *t* é um limiar de cotação que preserva proporções dentro da partição. Na Figura 7(*a*), a linha de cotações constantes é a bissetriz partindo, por exemplo, do vértice *A* que divide o 2 – *simplex* nas zonas $\frac{P(A)}{\max_p = P(B)}$ (acima da linha) e $\frac{P(A)}{\max_p = P(C)}$ (abaixo da linha). E isso era precisamente o que dizia a definição de *PL* para quando uma proposição deveria ser rejeitada do conjunto de crenças do agente.

No próximo capítulo iremos explorar uma das razões mais significativas para aderir *PL* como Princípio Ponte. Tal razão é devido ao fato de que o limiar de cotação funciona muito bem sob condicionalização bayesiana, num processo de revisão de crenças e probabilidade subjetiva, e essa é uma das razões para a teoria também ser conhecida por *teoria do rastreamento*. Mas isso é assunto para o próximo capítulo.

3.3.2 A geometria da Tese Humeana

Agora, para finalizar o capítulo, veremos como podemos visualizar a Tese Humeana de Leitgeb (2017) a partir de sua forma geométrica. Relembrando, enquanto o conjunto de crenças [**B**] é determinado em *TL* a partir de probabilidades absolutas e a partir de “cotações” em *PL*, $TH_{\mathcal{Y}}$ irá considerar probabilidades condicionais. Retomando o que diz $TH_{\mathcal{Y}}$, temos que *Bel* e *P* se relacionam da seguinte maneira: Para todo *Y* tal que $P(Y) > 0$, e $Y \in \mathcal{Y}$,

$$Bel(A) \text{ sse } P(A|Y) > \frac{1}{2}.$$

Como já vimos, um modo equivalente de compreender como $TH_{\mathcal{Y}}$ determina [**B**] é por meio da estabilidade dos núcleos de crença e pela condição de preponderância.

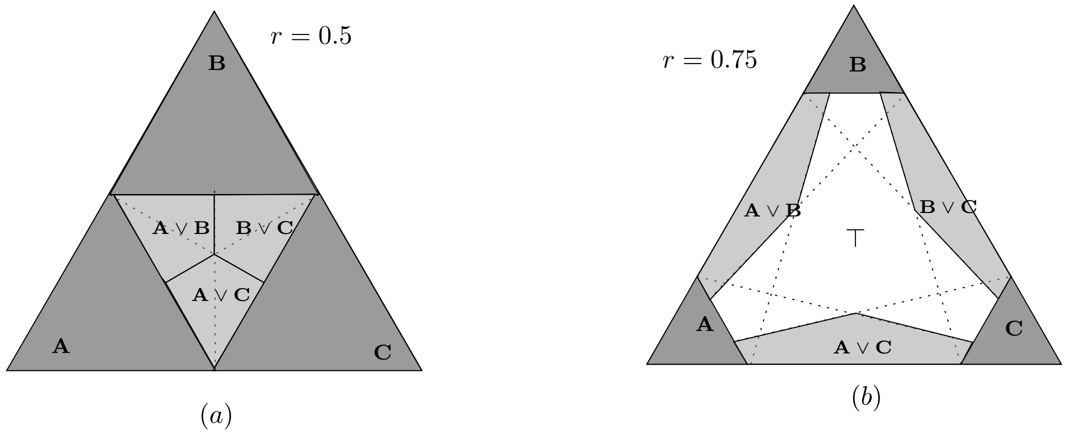

Figura 8 – A interpretação geométrica da Tese Humeana.

,

Esta, relembrando, satisfaz as seguintes condições:

1. para todo A, $Bel(A)$ *sse* $B_W \subseteq A$; e
2. B_W satisfaz satisfaz o seguinte (Condição de Preponderância): Para todo $w_i \in B_W$, é o caso que $P(w_i) > P(\neg B_W)$

Tal condição dá origem a forma geométrica de $TH_{\mathcal{Y}}$, como podemos ver na Figura 8(*a*). Ali podemos ver como o limiar $r = 1/2$ não gera qualquer região onde uma distribuição de probabilidade acarretaria um conjunto contraditório de crenças. Para visualizar como tal forma geométrica surge, basta olhar para a condição de preponderância a partir do modo como probabilidades são distribuídas em uma partição como um ponto no 2−*simplex*, tal como vimos na Figura 4. Assim, por sua forma geométrica, somos capazes de observar como $TH_{\mathcal{Y}}$ com limiar 1/2 é não-cético e consistente.

Já a Figura 8(*b*) nos oferece uma outra perspectiva. Até agora abordamos $TH_{\mathcal{Y}}$ a partir de um limiar $r = 1/2$. Mas como mostrou Leitgeb (2017), r pode assumir qualquer valor no intervalo (0.5, 1). A Figura 8(*b*) mostra como seria a estrutura geométrica de $TH_{\mathcal{Y}}$, com um limiar $r = 3/4$. Agora, para visualizar sua formação geométrica precisamos tomar a condição de preponderância como afirmando: Para todo $w_i \in B_W$, é o caso que $P(w_i) > \frac{r}{1-r} \cdot P(\neg B_W)$.

Como é possível observar, $TH_{\mathcal{Y}}^{3/4}$ (isto é, a Teoria Humeana com limiar 3/4) é muito mais cética que a *TL* sob um limiar equivalente, ou seja, onde $r = s$. Em outras palavras, é muito mais fácil uma determinada distribuição de probabilidade em $TH_{\mathcal{Y}}^{3/4}$ permitir apenas a crença na proposição tautológica $\top$, uma vez que a área branca do triângulo em $TH_{\mathcal{Y}}^{3/4}$ é muito maior que a área branca em *TL*. Essa é uma das razões pelas quais o PP humeano é consistente, enquanto o lockeano, não, uma vez que ele permite um número menor de proposições aceitas no conjunto de crenças do agente.

Por outro lado, podemos comparar $TH_{\mathcal{Y}}^{3/4}$ com *PL*, a partir das figuras 8(*b*) e 7(*b*) respectivamente. Como é possível observar, sob limiares equivalentes, $TH_{\mathcal{Y}}$ é mais cética que *PL*, uma vez que a área em branco no 2 – *simplex* é maior em 8(*b*) do que em 7(*b*). Além do fato de isso tornar uma teoria mais cética que a outra, essa pequena diferença geométrica será relevante no capítulo seguinte, onde veremos como tal diferença entre $TH_{\mathcal{Y}}^{3/4}$ e *PL* possui consequências fundacionais na forma como construímos uma lógica para a revisão (qualitativa) de crença.

Antes de encerrar este capítulo, gostaria de expandir um pouco mais o debate em torno da relação entre informatividade vs. ceticismo das opções de PP que foram apresentadas aqui. Relembrando, uma teoria será tão cética quanto ela torna raro o caso onde o agente termine com um conjunto informativo de crença, isto é, um conjunto que não seja meramente a disjunção de todas as opções disponíveis. Como foi possível ver, *TL*, $TH_{\mathcal{Y}}$, e *PL* geram uma hierarquia sobre o quão cético um agente deve ser, dado limiares equivalentes. Ou seja, *TL* é a menos cética dentre elas (a que possui menos área branca no 2 – *simplex*), seguido por *PL*, e finalizando com $TH_{\mathcal{Y}}$, esta, sendo a mais cética. Na medida em que *TL* se destaca como sendo a opção menos cética, ela permite que proposições sejam aceitas onde elas aparentemente não deveriam ser, e isso faz com que, como vimos, algumas distribuições de probabilidade permitam que o agente aceite proposições contraditórias. Por outro lado, enquanto *PL* e $TH_{\mathcal{Y}}$ são prudentes o suficiente para evitar a inconsistência de *TL*, elas divergem em muitos aspectos em sua capacidade de gerar conjuntos de crença informativos.

Em um louvável trabalho, Douven e Rott (2018) e Rott (2017) comparam as três supracitadas propostas de PP com um número maior de mundos e com os mais variados limiares. Os autores mostram como a chance[23] de um agente sob cada uma das regras possuir um conjunto de crenças não trivial é influenciado conforme o valor do limiar aumenta e também pelo número de mundos em consideração. Por exemplo, Rott (2017, p. 195) mostrou que, dado um limiar humeano de 0.9, a chance de haver uma distribuição de probabilidade que entregue um conjunto não trivial de crenças é de aproximadamente 18% quando $|W| = 3$, e em torno de 16% para $|W| = 4$. O principal responsável por esse resultado é o requisito de estabilidade para o conjunto de crenças em $TH_{\mathcal{Y}}$. Na medida em que o limiar humeano cresce, ou o número de mundos aumenta, conjuntos estáveis se tornam, nas palavras de David C Makinson (2015), cada vez mais "escassos". Por exemplo, no caso de um limiar de 0.8 e $|W| = 10$, a chance de *TL* entregar um conjunto de crenças informativo é 100%, para *PL* é de 99%, enquanto para $TH_{\mathcal{Y}}$ é de apenas 29% (DOUVEN; ROTT, 2018, p. 14). Isso mostra o quanto a Tese Humeana é mais cética em relação a suas rivais, principalmente sua

[23] "Chance" aqui não deve ser entendido no sentido estrito como vemos em parte da literatura bayesiana sobre probabilidades objetivas. A ideia aqui é simplesmente se referir a probabilidade de uma distribuição de probabilidade possível. No texto original os autores utilizam *"likelihood"*.

rival consistente, a Teoria Probalógica.

Apesar de $TH_{\mathcal{Y}}$ ser a opção mais cética, Leitgeb (2017, p. 127) argumenta que, dado um limiar de 1/2 para a Tese Humeana, "praticamente todas"[24] distribuições de probabilidade sobre W possuem pelo menos um conjunto P–estável com probabilidade menor que 1. E isso pode ser observado na Figura 8(*a*). Ainda, para o autor, em contextos ordinários é comum que deliberemos sobre uma determinada partição de "alternativas suficientemente prováveis", de modo que o número de mundos permaneça baixo o suficiente para tornar a teoria não-trivial (LEITGEB, 2017, p. 144).

Todavia, outro desafio enfrentado pela Tese Humeana é o que Schurz (2019) chamou de "problema da mente fechada": $TH_{\mathcal{Y}}$ só é satisfazível por pequenos conjuntos de possibilidades doxásticas. E uma consequência desse resultado é que quanto menor for o valor r para o limiar de $TH_{\mathcal{Y}}$, o que, conforme vimos acima, daria origem a um agente menos cético, menor será o número de proposições que o agente terá em seu conjunto *Poss* de possibilidades doxásticas. Abaixo, veremos brevemente como isso ocorre.

Retomemos o modo pelo qual o núcleo de crenças B_W é gerado. Dado um conjunto de mundos possíveis W, B_W representa a intersecção de todas proposições acreditadas pelo agente, de modo que $Bel(A)$ se, e somente se, para todo $w_i \in B_W$, temos que $w_i \in A$. Ou seja, o agente irá acreditar em uma proposição se ela for verdadeira em todos os mundos pertencentes ao seu núcleo de crenças. Por outro lado, o agente irá considerar uma proposição B como uma possibilidade doxástica, isto é, $Poss(B)$ se, e somente se, existe pelo menos um $w_j \in B_W$ onde $w_j \in B$. Com isso, uma consequência já explorada de $TH_{\mathcal{Y}}$ é que, pela Condição de Preponderância (conforme definição 3.1.2), toda possibilidade doxástica, ou seja, todo $w_i \in B_W$, é mais provável que a suposição de que B_W contenha um erro, ou seja, $P(\neg B_W)$.

Conforme mencionado acima, outra consequência para $TH_{\mathcal{Y}}$, de acordo com Schurz (2019), é que a teoria está comprometida com uma visão onde o número de possibilidades doxásticas para o agente é, em grande parte dos casos, demasiadamente pequeno, o que lhe renderia a alcunha de "mente fechada". O que Schurz (2019, p. 10) mostra é que o número de possibilidades doxásticas em $TH_{\mathcal{Y}}$ será sempre menor do que $\frac{P(B_W)}{1-r}$. Apresentaremos a seguir como esse resultado é obtido.

Seja $|B_W|$ o número de mundos pertencentes a B_W, e w_m o elemento de B_W com menor probabilidade (que existe uma vez que $|B_w|$ é finito). Pela Condição de

[24] Segundo o autor, "Praticamente todas" aqui pode ser entendido de modo mais preciso em termos de medidas de Lebesgue, em Teoria da Medida (para uma breve apresentação sobre o tema, ver (CORTZEN; WEISSTEIN, E. W., s.d.)). Como vimos na Figura 8(*a*), a única distribuição de probabilidade no 2 – *simplex* que implica em um conjunto cético de proposições aceitas é precisamente a distribuição uniforme, ou seja, onde todos os mundos possuem exatamente a mesma probabilidade. E essa distribuição é representada no triângulo por um ponto em seu centro. Como um ponto possui medida 0, praticamente todo ponto no 2 – *simplex*, ou seja, toda distribuição de probabilidade, irá resultar em um conjunto não trivial.

Preponderância, temos que $P(w_m) > 1 - r \geq P(\neg B_W)$. Dada uma distribuição de probabilidade qualquer sobre os elementos de B_W, sabemos que $P(w_m)$ será ao menos tão pequena quanto $\frac{P(B_W)}{|B_W|}$. Disso se segue que $\frac{P(B_W)}{|B_W|} > 1 - r$, e portanto, $|B_W| < \frac{P(B_W)}{1-r}$. O que isso significa para nós? Primeiramente, que quanto menor for o limiar para $TH_{\mathcal{Y}}$, menor será o valor possível para $|B_W|$, ou seja, menor será o número de mundos que o gente aceitará como possibilidades doxásticas. Vamos ver como isso se aplica ao exemplo da Tabela 2 apresentado aqui no capítulo.

Seja P uma distribuição de probabilidade sobre os 8 mundos representandos todas possíveis combinações para as proposições A, B e C, conforme a Tabela 2. Seja $r < P(w_1)$ de modo que $B_W = \{w_1\}$. Pelo resultado do parágrafo acima, $|B_W| < \frac{P(w_1)}{1-r} \approx 1,04$. Ou seja, como $|B_W|$ é sempre um número inteiro, B_W não poderá ter mais que 1 mundo. Nesse caso, temos que $[\mathbf{B}] = \mathcal{Y}$, ou seja, o agente acredita em tudo, e somente naquilo, que ele considera doxásticamente possível. Já no caso onde $r < 0,896$, de modo que $B_W = \{w_1, w_2, w_3, w_4\}$, temos que $|B_W| < \frac{P(B_W)}{1-r} \approx 8,6$. Ou seja, B_W não poderá ter mais que 8 mundos.

Com tais resultados podemos concluir algumas coisas sobre $TH_{\mathcal{Y}}$. Primeiramente, a partir de Douven e Rott (2018), conforme o limiar r cresce, ou o número de mundos em consideração aumenta, fica cada vez mais difícil para a Tese Humeana entregar um conjunto não-trivial de crenças. Ou seja, conjuntos de crenças informativos que satisfazem as condições exigidas por $TH_{\mathcal{Y}}$ são escassos. Uma solução para isso, conforme vimos com a resposta de Leitgeb, seria manter o número de mundos relativamente baixo (determinados por uma questão relevante) e o limiar não muito acima de 1/2. Contudo, o que Schurz (2019) mostrou é que, quanto menor for o valor do limiar, mais "mente fechada" se torna o agente, de modo que seu conjunto de crença seja muito parecido com seu conjunto de possibilidades doxásticas. Acredito que, tomadas em conjunto, essa é a mais dura objeção à teoria desenvolvida por Leitgeb, sendo assim um grande desafio ao nosso propósito de encontrar a melhor versão disponível para um princípio ponte que conecte crença e probabilidade subjetiva. Ainda assim, tanto a Tese Humeana quanto a Teoria Probalógica possuem seus méritos e dificuldades. No próximo capítulo vamos mudar um pouco a abordagem e explorar como tais teorias lidam com a dinâmica de revisão de crenças e probabilidades subjetivas com o passar do tempo frente novas evidências.

4 EM BUSCA DE UM PRINCIPIO PONTE DIACRÔNICO

> *Intending to have Cheerious for breakfast, Mary goes to the cupboard. But she can't find any Cheerios. She decides that Elizabeth must have finished off the Cheerios the day before. So, she settles for Rice Krispies. In the process, Mary has modified her original intentions and beliefs*. (Gilbert Harman, 1986, Capítulo I, p. 01)

Recapitulando brevemente nosso trajeto até aqui, iniciamos a presente Tese apresentando uma concepção de racionalidade capaz de abranger nossa visão teórica mais compartilhada sobre a racionalidade de dois tipos relevantes de atitudes doxásticas. Vimos que, dentre os requisitos da racionalidade, um deles era o alvo principal de nosso interesse: um requisito que conecte as duas noções doxásticas relevantes, a saber, crença e probabilidade subjetiva. Chamamos isso de uma busca por um "Princípio Ponte" (PP) que conecte a racionalidade das duas noções. Terminamos o primeiro capítulo apresentando o maior desafio para nosso projeto: a visão teórica mais intuitiva sobre como conectar a racionalidade de tais noções sofre de um problema de consistência.

No segundo capítulo demos início à investigação na busca por um PP apropriado a partir de uma visão *decisória* sobre a racionalidade da crença e da probabilidade subjetiva: crença (ou aceitação) é o resultado de um processo deliberativo que leva em conta a vontade de informar e o risco de errar. A partir das razões apresentadas no capítulo, argumentei em favor da ideia de que consistência e fechamento dedutivo devem permanecer sendo requisitos racionais da crença, e como consequência disso somos obrigados a aceitar um PP que determine a racionalidade da crença como dependente do contexto.

Por fim, no terceiro capítulo, nos dedicamos a analisar mais profundamente duas das principais propostas de PP que preservam fechamento dedutivo e consistência lógica: a *Tese Humeana* de Leitgeb (2017), e a *Teoria Probalógica* de Lin e Kelly (2012a,b). Nesse caso, foi discutida apenas a visão de racionalidade que toma estados doxásticos como estáticos: temos uma "foto" de nossa "caixinha" de crenças e olhamos para ela visando que ela seja logicamente consistente e dedutivamente fechada. Agora, no quarto e último capítulo, mudaremos um pouco a direção e olharemos para um dos elementos mais importantes em nossa fala sobre racionalidade: a capacidade de mudança de visão através do raciocínio[1].

A primeira parte deste capítulo é dedicada a uma discussão menos formal e mais epistemológica sobre a natureza da racionalidade diacrônica. Nele eu dou voz a uma discussão, pouco explorada na literatura até então, sobre os desafios em torno de uma concepção dinâmica da racionalidade, e ofereço uma proposta do que considero ser o modo mais apropriado de lidar com tais desafios. Considero esta uma de minhas contribuições mais significativas ao capítulo.

Com exceção do primeiro e último tópico, respectivamente, Seção 4.1 e 4.4, este capítulo será bastante técnico e formal. Para tornar a leitura mais agradável, a demonstração de alguns resultados é deixada para o Apêndice B, onde apresento os resultados já conhecidos na literatura de um modo mais introdutório e didático ao leitor não especializado.

4.1 O QUE É RACIONALIDADE DIACRÔNICA?

No primeiro capítulo, ao apresentar os requisitos iniciais básicos para crença e probabilidade subjetiva, mencionamos brevemente a existência de princípios *diacrônicos*. Um requisito racional diacrônico é um critério de racionalidade capaz de atuar na dinâmica temporal da racionalidade de estados doxásticos. Isso significa que, dado um momento t_1 anterior a t_2, o conjunto de estados doxásticos do agente em t_1 deve possuir alguma relação com o conjunto de estados em t_2, e tal relação pode ser avaliada de um ponto de vista da racionalidade do agente[2].

Ainda retomando o que foi apresentado no primeiro capítulo, ao apresentar os elementos básicos do bayesianismo, dissemos que o princípio de condicionalização é amplamente aceito por bayesianos como um requisito racional dinâmico para probabilidades subjetivas. Ainda, dissemos que o bayesianismo é constituído por três pilares: probabilidade subjetiva, teoria da probabilidade de Kolmogorov, e o princípio de condicionalização. Enquanto o primeiro dos pilares diz respeito a uma tese psicológica/metafísica, o segundo diz respeito à racionalidade. Mas sobre o que é o princípio de condicionalização? Ao longo deste tópico, pretendo encaminhar brevemente algumas considerações sobre isso, e veremos que a resposta não é tão trivial quanto querem muitos bayesianos. Farei isso partindo da tese mais comum sobre condicionalização: condicionalização é simplesmente mais um requisito da racionalidade, assim como seus irmãos sincrônicos. Pretendo mostrar que tal afirmação envolve fortes compromissos teóricos, aos quais não pretendemos aderir plenamente. Em segundo

[1] É importante reforçar que não estaremos preocupados aqui com a psicologia do raciocínio, mas sim com a estrutura de inferências e atualizações do conjunto de crenças e probabilidades subjetivas com o passar do tempo, frente novas evidências.

[2] Por simplicidades, e como de costume, "tempo" é modelado aqui como uma medida discreta ao invés de contínua.

lugar, pretendo oferecer algumas razões (até então originais na literatura) para tomar o princípio de condicionalização como simplesmente um guia para "mudança de visão".

4.1.1 Tese 1: Condicionalização é um tipo de requisito racional

Novamente, é importante relembrar aqui alguns elementos apresentados no primeiro capítulo, mais precisamente, Seção 1.2.2. Neste, apresentamos os requisitos racionais para probabilidade subjetiva, que se resumem aos axiomas da teoria da probabilidade. Aqui, chamo a atenção para o fato de que tais princípios oriundos da teoria da probabilidade são estáticos, isto é, eles avaliam apenas a racionalidade de um momento. Por exemplo, os axiomas da teoria da probabilidade dizem apenas sobre como deve ser meu conjunto de probabilidades subjetivas nesse exato momento, mas não são suficientes para explicar o que deverá acontecer com minhas probabilidades subjetivas ao longo do dia, conforme novas informações são adquiridas. Inclusive, o mesmo ocorre com a definição de probabilidade condicional. O que a probabilidade condicional exige é que o agente, em um determinado momento, tenha crenças condicionais que respeitem tal postulado: minha probabilidade subjetiva em uma proposição A deve ser recalculada, nesse mesmo instante, se estou assumindo, hipoteticamente, outro evento B como certo. Para passar da racionalidade de uma probabilidade subjetiva condicional para a racionalidade da resposta do agente frente a novas evidências recebidas no instante seguinte (ou seja, quando a condição inicialmente hipotética B "surge" como uma nova evidência), precisamos de um princípio genuinamente diacrônico.

Relembrando, no Capítulo 1, a Definição 6 estabelece o princípio de condicionalização como: para qualquer tempo t anterior a t', se o agente recebe a evidência, em um instante entre t e t', de que é o caso que B, e $P(B) \neq 0$, então $P_{t'}(A) = P_t(A|B)$. Note que o papel desempenhado pela condicionalização é precisamente de ser uma norma que conecta requisitos racionais ao longo do tempo. O problema com tal norma é o fato de que ela não possui qualquer semelhança com o que viemos assumindo até aqui como sendo requisitos para a racionalidade de um conjunto de estados doxásticos. Na concepção assumida desde o início da tese, racionalidade é sobre harmonia entre estados e atitudes doxásticas, não sobre harmonia temporal. Tal fato está refletido na dificuldade que bayesianos têm em justificar a racionalidade do princípio de condicionalização por meio do tradicional argumento *Dutch Book*, como mencionamos no primeiro capítulo, Seção 1.2. Essa dificuldade em justificar a existência de requisitos racionais diacrônicos aparece principalmente em Christensen (1991), Moss (2015), Arntzenius (2003) e Hedden (2015).

Um dos primeiros argumentos para justificar o princípio de condicionalização foi oferecido por David Lewis (1999), na forma de um argumento *Dutch Book*. O autor sustenta que, ao violar o princípio de condicionalização, o agente epistêmico torna-se susceptível à perda monetária em uma sequência de apostas ao longo do tempo.

Resumidamente, argumentos a favor da condicionalização dependem de que o agente tenha certo conhecimento sobre suas atitudes doxásticas com o passar do tempo[3]. Todavia, existe uma diferença significativa entre susceptibilidade de perda monetária instantânea e a susceptibilidade de perda monetária ao longo do tempo. Em outras palavras, existe uma diferença importante entre argumentos *Dutch Book* sincrônicos e argumentos *Dutch Book* diacrônicos. Tal diferença pode ser expressa por meio da simples pergunta: o que conecta a racionalidade de um conjunto de estados doxásticos de um agente em t_1 com a racionalidade do conjunto de estados do mesmo agente em t_2?

Alguns autores (ex. Briggs (2009) e Stalnaker (1984)) sustentam que uma possível resposta é exigir que o agente adote uma estratégia de planejamento para atualização frente novas evidências, e na medida em que tal planejamento é uma política temporal, a racionalidade do plano justificaria a racionalidade da condicionalização. Isso significa que se o plano estiver de acordo com a condicionalização, o agente será capaz de afirmar em t_1 que ele cumprirá ou violará condicionalização em t_2. Todavia, o problema com essa resposta é que condicionalização é uma norma diacrônica, para violá-la é necessário que o agente tenha um conjunto de probabilidades subjetivas em t_1 que não se adapte ao seu conjunto em t_2. Mas o argumento em defesa da condicionalização por meio de um plano é uma exigência apenas em t_1: um argumento sobre a adequação do conjunto de probabilidades subjetivas do agente em t_1 e seus planos e expectativas, no mesmo momento t_1, para atualização futuras. Tentarei tornar as coisas mais precisas com um exemplo.

Suponha que um agente possua um conjunto de probabilidades subjetivas em t_1 e um plano em consonância com a condicionalização em t_1. Todavia, no momento seguinte, em t_2, constatamos que ele falhou em cumprir com a condicionalização. Nesse caso, mesmo o agente descumprindo o princípio de condicionalização, nenhum argumento de aposta pode ser feito contra ele, uma vez que tal argumento seria baseado em seu "plano de condicionalização" em t_1. Por outro lado, um agente que não possui um plano de atualização em acordo com a condicionalização, em t_1, pode ser acusado de irracionalidade mesmo que em t_2 ele venha a cumprir as exigências estabelecidas pela condicionalização. O desafio central é o seguinte: Como podemos argumentar que a racionalidade exige que o agente honre com seus compromissos anteriores em um momento seguinte?

Argumentando contra a condicionalização como um requisito racional diacrônico, Christensen (1991) apresenta o exemplo de um apostador clarividente. Tal apostador tem a capacidade de determinar não somente os graus de crença do agente no presente momento, mas também seu grau de crença daqui a uma hora. Vamos supor que

[3] Aprofundar os detalhes sobre tal discussão desviaria demasiadamente o nosso foco. Para o leitor interessado, Titelbaum (2022b), nos capítulos 9 e 10, fornece uma apresentação rica em detalhes sobre esse desafio.

o agente altere sua distribuição de probabilidades subjetivas nesse tempo. O apostador, prevendo isso, pode facilmente assegurar seu ganho oferecendo uma aposta com cotações apropriadas. Para não ser susceptível a tal golpe, parece-nos que o agente deveria satisfazer um requisito racional diacrônico para probabilidades subjetivas bastante forte, uma condição que proibiria qualquer mudança nos graus de crença com o passar do tempo. Nas palavras do autor, essa exigência seria um tipo de "calcificação" dos estados doxásticos do agente. Mas tal exemplo não parece proporcionar qualquer razão para tomar a "calcificação" como um requisito racional. Um agente epistêmico que esquece uma de suas crenças não deveria ser considerado irracional. Memória e racionalidade parecem faculdades distintas. Segundo Christensen (1991, p. 241), as crenças de um agente em momentos distintos não exigem o mesmo tipo de consistência que o conjunto de crenças momentâneo.

Ainda podemos expandir os contra-exemplos para casos ainda mais curiosos. Como bem apontou Christensen (1991, p. 239), um sagaz apostador profissional pode realizar uma conjunto de apostas enviesadas (um *Dutch Book*) contra um casal, e facilmente garantir seu lucro, não importa o resultado. Mas isso não significa que cada membro do casal, tomado individualmente, está sendo irracional. Em outras palavras, é possível que, em conjunto com outras pessoas, nossas probabilidades subjetivas não satisfaçam os axiomas probabilísticos, ao mesmo tempo que não faz sentido dizer que cada um de nós está sendo individualmente irracional. Nas palavras do autor, "[c]onsistência em graus de crença, assim como consistência dedutiva, é um ideal racional para indivíduos, não para casais – mesmo para um casal com conta conjunta" (CHRISTENSEN, 1991, p. 240).

Seguindo essa ideia, exigir que um mesmo agente epistêmico possua uma unidade temporal em seu conjunto de estados doxásticos se assemelha com a exigência de que diferentes indivíduos possuam uma unidade coletiva entre seus estados. Em outras palavras, não somos capazes de explicar porque requisitos racionais temporais são mais "importantes" que requisitos racionais interpessoais. E não é difícil perceber que isso é o caso ao menos do ponto de vista de nossos critérios para avaliação de racionalidade estrutural. Nada do que foi apresentado e assumido aqui sobre racionalidade, até então, nos auxilia a distinguir entre esses dois casos. No máximo, podemos garantir que um argumento de aposta ou acurácia pode estabelecer que é racional para o agente planejar, ou ter expectativas, de acordo com a condicionalização, o que não é suficiente para a existência de um requisito genuinamente diacrônico.

Tal problema na concepção de racionalidade diacrônica está refletido na tese da "Racionalidade de Fatias Temporais" (do inglês *Time-slice Rationality*), presente nos trabalhos de Sarah Moss (2015) e Brian Hedden (2015). Segundo essa tese, um defensor da racionalidade de fatias temporais sustenta que requisitos racionais são inteiramente determinados pelos estados mentais daquela fatia temporal, e fatos fundamentais sobre racionalidade são esgotados por fatos sobre a racionalidade de fatias temporais. Essa visão substitui requisitos racionais diacrônicos por normas sincrônicas, onde tais normas dizem o que é racional para determinadas fatias temporais. Para Moss (2015, p. 172), podem existir fatos sobre o que é exigido de um agente racional estendido temporalmente, mas tais fatos são deriváveis de requisitos sincrônicos, e um agente poderá ser considerado racional apenas na medida em que é constituído de fatias temporais racionais, ou que ao menos grande parte de suas fatias temporais sejam racionais.

Além disso, segundo Hedden (2015, p. 452), a tese da racionalidade de fatias temporais está comprometida como uma afirmação de que requisitos racionais são impessoais: racionalidade é fundada em fatos sobre relações normativas que sobrevêm a um conjunto de estados mentais. Em outras palavras, o que você deve acreditar não depende de quem você é, ou seja, não depende de uma tese metafísica sobre individualidade. Desse modo, estabelecer em que um agente racional deve crer não exige fatos sobre identidade pessoal, ou uma visão de agente em momentos distintos. E não é difícil notar que tal concepção parece bastante harmônica com a visão de racionalidade estrutural que viemos apresentando desde o início desta tese. Então, o que devemos fazer com o princípio de condicionalização?

4.1.2 Tese 2: Condicionalização é sobre mudança de visão.

Frente aos desafios apresentados acima, considero aqui que uma possível interpretação para o princípio de condicionalização é tratá-lo não como um requisito estrutural da racionalidade, mas como um guia para "mudança de visão", ou também "mudança de visão raciocinada"[4]. Isso é inspirado fundamentalmente pelos *insights* de Gilbert Harman em sua obra *Changing in View* (1986). Logo na abertura, Harman apresenta um exemplo semelhante ao seguinte:

> Em uma manhã típica, Mary acorda e caminha em direção à cozinha para seu café da manhã. Mary pretende, como de costume, tomar seu leite com cereais. Mas, inesperadamente, ela não encontra os cereais. Mary descobre que Elizabeth, sua colega de quarto, terminou com os cereais no

[4] "Mudança de visão" pode ser entendido aqui como "raciocínio". Contudo, opto pelo primeiro em razão de o termo raciocínio ser geral o suficiente para incluir casos de inferência a partir do próprio conjunto inicial de estados, enquanto "mudança de visão" se refere exclusivamente ao processo dinâmico de atualização do conjunto doxástico, dado novos *inputs*. Raciocínio sem mudança de visão poderia ser chamado aqui de "reflexão", uma vez que só depende do agente realizando inferências a partir de seu conjunto inicial, onde tal "base" inicial, a partir da qual o agente raciocina, permanece estática, apesar de sua "reflexão" também ser algo que acontece de modo dinâmico ao longo do tempo. Tal concepção de raciocínio não terá relevância aqui, uma vez que, desde o início, estamos assumindo um agente epistêmico inferencialmente perfeito, e isso implica que ele já acredita em todas as consequências lógicas de seu conjunto de crenças. Para uma visão semelhante, ver Rott (2001, cap. 3).

> dia anterior. Desse modo, Mary decide por abrir um novo pacote de sua granola predileta para substituir os cereais.

Segundo Harman, nesse processo, Mary modificou sua intenção e crença original. Como nosso interesse é epistêmico, e não prático, daqui em diante vamos nos dedicar somente ao caso de alteração no conjunto de crenças.

Conforme o exemplo acima, o elemento central para mudança de visão é a existência de uma perturbação externa: dado um conjunto de atitudes doxásticas em equilíbrio, ou seja, satisfazendo requisitos racionais estruturais sincrônicos, estamos interessados no que acontece com tal conjunto quando ele permite novos *inputs* ao longo do tempo. Em resumo: toda mudança de visão resulta de um *input*. Tais *inputs* podem vir de duas formas: o agente pode receber um conteúdo novo ou ficar em dúvida sobre algo que ele já acreditava ser verdadeiro. Para Harman (1986, p. 02), mudança de visão envolve *Máximas de revisão*, que dizem respeito à mudança momentânea que deve ser feita, ou seja, a mudança que é parte do processo de revisão. Tal tipo de máxima é justamente uma norma diacrônica de revisão ao longo do tempo, como é o caso do princípio de condicionalização bayesiana. Contudo, o tópico anterior mostrou as dificuldades justamente de se aceitar esse tipo de máxima como um tipo de requisito racional. Então voltamos exatamente ao ponto zero, com a seguinte questão: como justificar a racionalidade de um princípio realmente diacrônico?

Não é difícil perceber que nossa dificuldade é fruto precisamente da nossa incapacidade de justificar o requisito de "Persistência da Crença", apresentado aqui no primeiro capítulo, Seção 1.2.1. Resumidamente, o que tal norma exige da racionalidade é que o agente não esqueça ou abandone suas crenças aleatoriamente. Isso significa que o requisito de Persistência da Crença constrói uma ponte entre dois momentos distintos, dentro do mesmo conjunto de estados, de um mesmo agente epistêmico. Contudo, nosso problema continua exatamente o mesmo do tópico anterior: parece que memória e racionalidade são elementos distintos; falhas de memória não parecem equivalentes a falhas da racionalidade.

Apesar de, ao menos em um primeiro momento, não sermos capazes de justificar a racionalidade do requisito de *Persistência da Crença*, ele parece uma condição fundamental se queremos falar em raciocínio ou mudança de visão. Olhando mais atentamente, o requisito de *Persistência da Crença* é estruturalmente distinto do requisito de condicionalização: enquanto a racionalidade do segundo parece ter origem na racionalidade dos requisitos sincrônicos para probabilidade subjetiva, a racionalidade do primeiro exige uma justificação própria, isto é, *Persistência da Crença* parece uma norma fundamental, não derivável, para racionalidade diacrônica. Sendo assim, se assumimos o requisito de *Persistência*, crenças serão estáveis ao longo do tempo, e o princípio de condicionalização adquire a estabilidade temporal que falta nos argumentos em sua defesa, como vimos acima. E como consequência, a racionalidade da condicionalização pode agora ser justificada por um argumento ao estilo *Dutch Book*, mas agora com a garantia de que as probabilidades subjetivas permanecerão "calcificadas" (usando as palavras de Christensen). Como podemos notar, o que há de genuinamente diacrônico no requisito de condicionalização é sua subjacente suposição da racionalidade da *Persistência* aplicada à probabilidade subjetiva. Isso quer dizer que a condicionalização deve ser entendida apenas como uma ferramenta prática para o agente satisfazer os requisitos racionais sincrônicos, na medida em que novas evidências são adquiridas ao longo do tempo, e o agente deverá ajustar seu novo conjunto de crenças à nova evidência adquirida para satisfazer os requisitos sincrônicos. Ou seja, a racionalidade do requisito bayesiano de condicionalização é fruto dos requisitos racionais sincrônicos para probabilidade subjetiva somado ao requisito de *Persistência*.

O que podemos afirmar até então é que mudança de visão é algo essencialmente dinâmico, e também é bastante plausível sustentar que a mudança de visão é algo que pode ser feito de modo mais ou menos racional. Isso significa que parece haver alguma relação entre os fenômenos referidos pelos conceitos de racionalidade e mudança de visão. Mas ao abrir mão da tese de que condicionalização é um requisito racional, não somos capazes de dizer precisamente qual a relação entre a racionalidade dos requisitos racionais sincrônicos (os axiomas probabilísticos) e o princípio de condicionalização. Minha resposta ao desafio sobre a conexão entre racionalidade e mudança de visão é a seguinte: devemos tomar o segundo como uma ferramenta para o primeiro. Na medida em que racionalidade é passiva, mudança de visão é ativa. Uma inspiração semelhante pode ser encontrada no trabalho de Broome (2013). Segundo o autor, "[Mudança de visão] é algo que fazemos que pode nos levar a satisfazer requisitos da racionalidade"[5] (BROOME, 2013, p. 207). Como veremos no próximo tópico, a dinâmica da crença só se torna um problema interessante na medida em que exigimos consistência e fechamento dedutivo como um ponto de partida, ou seja, se partirmos dos requisitos racionais sincrônicos para crença qualitativa. A adição de novas crenças pode levar a violação do ideal de consistência, enquanto a remoção de crenças pode violar o ideal de fechamento dedutivo. Se tais requisitos sincrônicos são abandonados, fazer atualização de crenças é simplesmente adicionar e remover os novos *inputs*.

Com ou sem a garantia da existência de um requisito racional como o de *Persistência da Crença*, estudar processos de atualização por meio de mudança de visão parece-nos algo relevante. E isso é essencialmente um processo dinâmico que depende de requisitos racionais estáticos. Frente à dificuldade de se justificar a racionalidade de uma norma essencialmente dinâmica, considero que uma possível solução seria adotar o requisito de *Persistência da Crença* como parte da definição de mudança

[5] O autor utiliza o termo "raciocínio" ao invés de mudança de visão. Opto pelo segundo pelas razões apresentadas na nota 4.

de visão, assim como fizemos no primeiro capítulo, Seção 1.2.3, com a definição de racionalidade epistêmica. Relembrando o que foi feito com a noção de racionalidade epistêmica no primeiro capítulo, dada a dificuldade em justificar a conexão entre crença e verdade, utilizamos tal conexão como uma forma de definir racionalidade epistêmica, em oposição à racionalidade prática. E fizemos isso sem apelar para algum tipo de norma fundamental (primitiva) da racionalidade. Assim, vejo essa como uma possível alternativa para o caso da racionalidade diacrônica, uma vez que é de nosso interesse investigar a racionalidade doxástica de processos e sistemas dinâmicos, e isso exige que ao menos postulemos alguma norma diacrônica (assumindo nossa dificuldade de justificá-la). Ao mesmo tempo, não queremos expandir a ontologia de requisitos da racionalidade para incluir algo como o princípio de *Persistência da Crença*, ou qualquer tese mais robusta sobre identidade pessoal temporal. Assim, pressupor a *Persistência* como um princípio racional para crença se faz necessário uma vez que, ao abrir mão da existência de qualquer norma diacrônica para a racionalidade, não somos capazes de investigar a mudança de visão como um fenômeno relevante para a racionalidade. E o objetivo deste capítulo é justamente expandir nossa discussão para casos onde há revisão temporal em estados doxásticos, e esse parece um objetivo digno de atenção[6].

Finalmente, agora sendo mais objetivo, podemos responder ao desafio central com o qual iniciamos este capítulo dizendo que a racionalidade de um requisito racional diacrônico (como o de condicionalização) é derivada da racionalidade dos requisitos sincrônicos para probabilidade subjetiva, somado a (suposição de) racionalidade da *Persistência da Crença*. Se consideramos importante o objeto de estudo da literatura que será explorada no restante do capítulo, e se temos a pretensão de explicar esse objeto por meio de elementos da racionalidade, é necessário, no mínimo, pressupor que *Persistência da Crença* é essencial para a mudança de visão racional. Além disso, como veremos no próximo tópico, teorias lógicas para revisão de crença também assumem alguns elementos não-lógicos (ou não-estruturais) em seus desenvolvimentos. Nesses casos, a justificação de tais princípios dependerá dos papéis desempenhados dentro da teoria, de certo consenso sobre o caráter intuitivo de tais pressupostos, e também sobre a necessidade de tê-los ao se fazer teoria de revisão de crença.

No próximo tópico veremos como o fenômeno de mudança de visão tomou a atenção de uma comunidade filosófica interessada em teorias formais sobre revisão de crença qualitativa. Como já foi dito aqui algumas vezes, enquanto o princípio de condicionalização é consensualmente a norma de revisão para probabilidades subjeti-

[6] Uma das considerações que torna tal objeto de estudo relevante está refletida no fato de que o estudo sobre mudança de visão é tema central aos teóricos da Inteligência Artificial. Nas palavras de Reiter (1987, p. 147), “Se os pesquisadores em Inteligência Artificial podem concordar em algo, é que um artefato inteligente deve ser capaz de raciocinar sobre o mundo em que habita. Esse artefato deve possuir várias formas de conhecimento e crença sobre esse mundo, e deve utilizar tal informação para inferir outras informações sobre o mundo visando tomar decisões, planejar e executar ações, responder outros agentes, etc.”

vas no bayesianismo, do lado qualitativo, não há algo próximo de um consenso sobre uma teoria para revisão de crenças. Vamos explorar a seguir a origem e possíveis prognósticos para esse debate.

4.2 UM MODELO DE MUDANÇA DE VISÃO PARA CRENÇA

Como vimos logo acima, o objetivo da mudança de visão é restabelecer harmonia no sistema doxástico frente a uma perturbação externa, principalmente quando tal perturbação ameaça a consistência do sistema. E é justamente a "inconsistência" no conjunto de crenças que nos parece uma boa razão para alguém realizar uma mudança de visão. Assim, o que regras diacrônicas devem visar, em um primeiro instante, é a satisfação dos requisitos racionais estáticos que garantem consistência. Todavia, como veremos, o maior desafio de uma teoria de revisão para crença qualitativa é lidar com o fato de que tal harmonia pode ser reestabelecida de diversas maneiras que são equivalentes de um ponto de vista da pura e simples consistência.

Para inciar nossa investigação em direção a uma lógica da mudança de visão para crença, utilizaremos de um exemplo semelhante ao que aparece em Rott (2001, p. 71):

> "Henry" é o nome dado por caçadores ao pássaro capturado na última tarde em uma armadilha em Varsóvia, capital da Polônia. Henry se parece exatamente com um cisne. Um amigo ornitologista me contou que todos cisnes europeus são brancos. Podemos resumir o conjunto pertinente de crenças nesse caso pela seguinte sequência de proposições:
>
> C: Henry é um cisne;
>
> P: Henry é Polonês;
>
> E: Henry é europeu;
>
> $C \wedge E$ acarreta B;
>
> B: Henry é branco.
>
> Todavia, ao ir verificar empiricamente o animal capturado, descobrimos para nossa surpresa que Henry é preto, e não um cisne branco!

Para explorar o que acontece no exemplo acima, o primeiro passo é notar que nosso conjunto de crenças recebeu uma nova informação, que é tomada como certa: (A) Henry é preto! Queremos que nosso conjunto de crenças atual contenha a nova proposição A. Todavia, uma vez que A acarreta $\neg B$, ao simplesmente adicionar A ao conjunto anterior, violamos o requisito sincrônico de consistência para o conjunto de crenças. Nas palavras de Hans Rott,

> "*Qualquer* tipo de requisito estático de coerência para conjunto de crenças dificulta as operações de mudança de crença. Mudança de crença sem restrições [...] pode ser satisfeita por operações diretas de *add-and-delete* em teoria dos conjuntos. Isso não é mais possível se é exigido de um conjunto de crenças que resulte em um conjunto de estados coerentes."(ROTT, 2001, p. 72)

Portanto, se pretendemos preservar a racionalidade estática do conjunto, precisamos *revisá-lo*. Isso significa que alguma das crenças do conjunto inicial deve ser removida, ou seja, utilizando um termo comum à literatura, revisão de crenças envolve *contração* de proposições. Como veremos na Seção 4.2.2, uma consequência disso é que a Lógica por trás de um modelo formal para mudança de visão precisa ser uma lógica *não-monotônica*.

O problema é que não queremos abandonar mais informações do que o necessário para reestabelecer a consistência do conjunto, e existe mais de uma opção de proposição a ser removida. Em nosso exemplo, podemos satisfazer o requisito de consistência removendo C, P, ou que $C \wedge E$ acarreta B. O mais importante aqui é o fato de que, para acomodar A no conjunto inicial, não existe qualquer elemento puramente lógico capaz de nos dizer qual das proposições acima deve ser removida do conjunto.

Além disso, o leitor pode ter notado no parágrafo anterior uma suposição extralógica de fundo: "não queremos abandonar mais informações do que o necessário". Tal afirmação está conectada com o famoso princípio de *Economia Informacional*. Isso significa que, do ponto de vista de uma teoria para revisão de crença, o conjunto inicial de crenças deve ser tomado como certo, na medida em que preserva o sistema em harmonia. E isso se mantém até que uma nova evidência entre no sistema e exija uma revisão. Tal doutrina é conhecida na literatura epistêmica como "conservadorismo epistemológico", e pode ser encontrada em autores como Harman (1986) e Quine e Ullian (1978).

Para acomodar esses elementos não-lógicos mencionados, em uma teoria da revisão de crenças precisamos de alguns postulados essencialmente diacrônicos. Segundo Rott (2001, p. 73), a remoção de crenças de um conjunto durante uma revisão deve satisfazer duas máximas:

i) a quantidade de informação perdida em uma mudança de crença deve ser mínima;

ii) na medida em que crenças são consideradas umas mais importantes que outras, o agente deve remover as menos importantes.

E, como veremos, é precisamente a existência de tais elementos não-lógicos que faz de uma teoria da revisão de crenças uma fonte interminável de controvérsias.

Nas próximas seções veremos duas teorias sobre qual o modelo mais apropriado para fazer revisão de crenças, começando pela teoria AGM, e posteriormente para um Sistema Preferencial não-monotônico. Para ambas as teorias, dado um conjunto inicial de crenças e um novo *input*, na forma de uma nova informação adquirida, existem tradicionalmente três formas de proceder: podemos simplesmente adicionar a nova informação adquirida no conjunto de proposições aceitas; pode ser que a nova informação diga apenas que uma de nossas crenças antigas deve ser abandonada; ou pode ser que ao adicionar a nova informação no conjunto de crenças, isso viole o requisito de consistência, e para nos manter consistentes seja necessário remover alguma das crenças antigas para acomodar a nova informação. Como veremos a seguir, essas três formas de proceder em uma mudança de visão são tradicionalmente conhecidas, respectivamente, por *Expansão, Contração*, e *Revisão*. É importante ressaltar que, enquanto o requisito sincrônico de consistência coloca uma restrição para as revisões, a exigência de fechamento dedutivo para o conjunto de crenças coloca uma restrição para as contrações. Sendo mais preciso, ao adicionar novas informações ao conjunto inicial de crenças, o requisito de consistência exige que qualquer proposição do conjunto inicial que entre em conflito com a nova informação deve ser removida. Por outro lado, ao remover uma informação do conjunto inicial de crenças, o requisito de fechamento dedutivo exige que todas as consequências lógicas da proposição removida sejam também abandonadas.

4.2.1 A Teoria AGM

A teoria AGM é reconhecida como o ponto de partida para aqueles que pretendem trabalhar com uma representação formal da revisão de crença. O trabalho fundante da teoria consiste em um artigo de 1985 escrito pelo trio Carlos Alchourrón, Peter Gärdenfors, e David Makinson (AGM é simplesmente as iniciais dos respectivos sobrenomes), intitulado *"On the logic of theory change: Partial meet contraction and revision functions"*. Antes de adentrar propriamente nos conceitos centrais de expansão, contração e revisão, faremos algumas considerações sobre nossa abordagem formal da teoria.

Seguindo a referida literatura em questão, o modo padrão de modelar um conjunto de crenças em uma teoria da revisão é por meio de conjuntos de *sentenças*. Diferentemente do modo como viemos trabalhando até então, com conjuntos de crença formados por proposições, neste tópico tal conjunto será precisamente aquele contendo as sentenças aceitas pelo agente, e isso exige a introdução de um vocabulário ainda não apresentado. Esse tipo de modelagem com sentenças pressupõe uma linguagem objeto $\mathcal{L}$, da qual as sentenças do conjunto de crenças serão selecionadas, onde letras minúsculas p, q, r denotarão sentenças atômicas particulares (ou constantes) em $\mathcal{L}$, α, β, γ denotarão sentenças quaisquer (metavariáveis para fórmulas-bem-formadas), enquanto Γ, Δ denotarão conjuntos de fórmulas. O conjunto especial de sentenças que corresponde ao nosso conjunto de crenças será representado por **B**, o qual cumpre os postulados sincrônicos para crença apresentados no primeiro capítulo[7].

Continuando, dizemos que o agente acredita em α se, e somente se, $\alpha \in \mathbf{B}$. Assumindo fechamento lógico para crença, temos que **B** é igual ao conjunto *Cn*(**B**), onde *Cn* é um *operador de consequência* tarskiano. Em uma notação alternativa, dizemos que $\alpha \in Cn(\mathbf{B})$ *sse* $\mathbf{B} \vdash \alpha$. Ainda, dizemos que **B** é um conjunto *não-trivial* de crenças se, e somente se, existe um $\alpha \in \mathcal{L}$ tal que $\alpha \notin \mathbf{B}$. Ainda, o conjunto de todas as sentenças logicamente válidas, ou seja, as sentenças que estão inclusas em todo conjunto de crenças, será denotado por $Cn(\emptyset)$. Por outro lado, o conjunto $\mathcal{L}$ de todas as sentenças é o conjunto absurdo, e será denotado por $\mathbf{B}_{\perp}$. Tecnicamente, podemos resumir isso tudo dizendo simplesmente que um conjunto de crenças contendo sentenças é o que os lógicos normalmente chamam de uma *Teoria*.

Estabelecido o mínimo vocabulário lógico necessário, nosso objetivo agora será apresentar, resumidamente, como a teoria AGM desenvolve os conceitos relevantes de expansão, contração e revisão. Boa parte do que será brevemente apresentado a seguir é baseado, fundamentalmente, no excelente trabalho de Gärdenfors (1988), e também auxiliado pelos textos de Hansson (2018, 2022)[8].

Começando pelo modo mais simples, temos o caso quando uma nova informação é adquirida, isto é, quando o agente aprende algo novo, e tal informação não entra em conflito com qualquer uma das crenças anteriores do conjunto do agente. Relembrando, chamamos esse caso de *Expansão*. Formalmente, chamamos de "+" uma função de expansão que mapeia o par $\mathbf{B} \times \mathcal{L}$ em direção ao conjunto de crenças $\mathbf{B}'$, onde $\mathbf{B}'$ pode ou não ser igual a **B**. A expansão de **B** pela sentença α é denotada como $\mathbf{B}^{+}_{\alpha}$. Uma vez que a informação adquirida não afeta a consistência do conjunto inicial de crenças, realizar uma expansão é simplesmente adicionar a nova informação ao conjunto, e realizar o fechamento lógico do novo conjunto, ou seja, $\mathbf{B}^{+}_{\alpha} = Cn(\mathbf{B} \cup \{\alpha\})$.

Dizemos que um processo de mudança de visão por meio de expansão é unicamente determinado, ou seja, dada uma nova informação α, só existe uma opção para o conjunto $\mathbf{B}^{+}_{\alpha}$. Isso quer dizer que expansão pode ser definida exclusivamente por operações tradicionais em teoria dos conjuntos. Todavia, isso não é possível no caso de uma revisão ou contração. Como vimos acima no exemplo do cisne Henry, existem vários modos de se remover uma sentença de um conjunto de crenças. Uma vez que revisão também envolve a remoção de alguma sentença do conjunto de cren-

[7] Para ser mais fiel às teorias que serão apresentadas aqui, trabalharemos com sentenças ao invés de proposições, e com **B** ao invés de [**B**] para representar o conjunto de crenças do agente epistêmico. Enquanto [**B**] é um conjunto de crenças contendo proposições, **B** é um conjunto de crenças contendo sentenças. No tópico 4.2.3 veremos como podemos realizar uma tradução entre ambas visões, e veremos que os requisitos para crença apresentados no primeiro capítulo podem ser formulados tanto sobre conjunto de proposições como conjunto de sentenças.

[8] Embora o que será apresentado aqui deixe de lado alguns detalhes teóricos relevantes para um entendimento mais avançado da teoria, a presente apresentação da teoria AGM é suficiente para o propósito do capítulo, e ainda, acredito que ela forneça uma boa introdução em nosso idioma, o que é um material bastante escasso.

ças, tanto a contração quanto a revisão envolvem certas *escolhas* sobre o que deve ser removido. A seguir, nos dedicaremos a explicar os possíveis modos de contrair e revisar sentenças de um conjunto de crenças.

Uma vez que o processo de remoção de sentenças de um conjunto não pode ser determinado exclusivamente por operações em conjuntos, é necessário que se construa uma função de contração que determine unicamente o conjunto final de crenças, dada uma nova sentença adquirida. O que os precursores da teoria AGM observaram é que, no processo de remoção de uma informação α, existiam vários subconjuntos de **B** que não acarretam α. Além disso, podem existir vários subconjuntos maximais, ou seja, conjuntos contendo o maior número de sentenças, que não acarretam α. Assim, seja $\div$ uma função de contração que mapeia o par $\mathbf{B} \times \mathcal{L}$ em direção ao conjunto de crenças $\mathbf{B}'$. A contração de **B** por uma sentença α qualquer é denotada por $\mathbf{B}^{\div}_{\alpha}$.

Nosso desafio então é escolher entre diferentes opções de conjuntos que representem $\mathbf{B}^{\div}_{\alpha}$ de tal modo que a função $\div$ esteja bem definida (uma vez que uma função não pode levar o mesmo argumento em dois elementos distintos do contradomínio). Dado o critério de economia informacional, visto no início desta seção, não queremos que nenhuma informação seja perdida desnecessariamente durante a contração. Uma ideia razoavelmente intuitiva é exigir que $\mathbf{B}^{\div}_{\alpha}$ seja o maior subconjunto possível de **B** que não acarrete α. Para isso, precisaremos introduzir um novo conceito. Chamemos de "*sobra*$_{\alpha}$ de **B**" (do inglês *remainder*) a coleção de todos subconjuntos de maximais de **B** que não acarretam α, e o denotaremos por $\mathbf{B}\perp\alpha$.

Seguindo Gärdenfors (1988, p. 77), a primeira ideia de construção para $\mathbf{B}^{\div}_{\alpha}$ é tomar uma *função de seleção s* que pega apenas um elemento de $\mathbf{B}\perp\alpha$. Tal construção é conhecida como "contração de escolha-máxima" (do inglês *maxichoice contraction*), e é definida por

$$\mathbf{B}^{\div}_{\alpha} = \begin{cases} s(\mathbf{B}\perp\alpha), & \text{quando não temos} \vdash \alpha \\ \mathbf{B}, & \text{quando} \vdash \alpha \end{cases}$$

O problema é que tal construção para contração possui propriedades indesejáveis. Como mostrou Gärdenfors (1988, p. 78), se $\alpha \in \mathbf{B}$ e definirmos $\mathbf{B}^{\div}_{\alpha}$ como uma contração de escolha-máxima, para qualquer outra sentença β, temos que ou $(\alpha \vee \beta) \in \mathbf{B}^{\div}_{\alpha}$ ou $(\alpha \vee \neg\beta) \in \mathbf{B}^{\div}_{\alpha}$. Desse modo, contrações de escolha-máxima geram grandes conjuntos de crença, isto é, todo elemento de $\mathcal{L}$ pertencerá à **B**. E uma vez que operações de revisão podem ser construídas por meio de contrações e expansões (como veremos mais abaixo), contração de escolha-máxima *cria* conjuntos de crenças maximais, o que não seria uma propriedade desejável. Desse modo, devemos buscar outra maneira de construir uma operação de contração.

Uma segunda opção é assumir que $\mathbf{B}^{\div}_{\alpha}$ contém somente proposições que são comuns a todos elementos de $\mathbf{B}\perp\alpha$. Para Gärdenfors (1988, p. 78), essa seria uma "contração de intersecção completa" (do inglês *full meet contraction*), e seria definida da seguinte maneira:

$$\mathbf{B}\dot{-}_{\alpha} = \begin{cases} \bigcap(\mathbf{B}\bot\alpha), & \text{quando } (\mathbf{B}\bot\alpha) \neq \emptyset \\ \mathbf{B}, & \text{quando } (\mathbf{B}\bot\alpha) = \emptyset \end{cases}$$

Isso significa que uma determinada informação β pertence ao conjunto $\mathbf{B}\dot{-}_{\alpha}$ se, e somente se, β está contida em *todos* os subconjuntos maximais de **B** que falham em acarretar α.

Assim como o caso anterior da contração de escolha-máxima, a contração de intersecção completa também possui características indesejáveis. Mas, enquanto a primeira pecava por gerar conjuntos de crenças muito grandes, a segunda peca por permitir apenas conjuntos muito pequenos. Resumidamente, seguindo Gärdenfors (1988, p. 79), ao utilizar intersecção completa para gerar um conjunto particular $\mathbf{B}\dot{-}_{p}$, somos conduzidos a um conjunto de crenças que contém somente as consequências lógicas de $\neg p$. Todavia, apesar de se tratar de uma opção limitada para definir a operação de contração, a contração de intersecção completa funciona como um *limitante inferior* para operações de contração. Isso quer dizer que qualquer conjunto $\mathbf{B}\dot{-}_{\alpha}$ resultante de uma contração deve incluir o conjunto gerado por $\bigcap(\mathbf{B}\bot\alpha)$.

Uma vez que as duas tentativas anteriores de definir uma função de contração resultaram em extremos opostos, uma opção natural é investigar a possibilidade de encontrar um meio termo entre elas. Dado que a contração de intersecção completa tomava a intersecção de *todos* os subconjuntos de $\mathbf{B}\bot\alpha$, uma possibilidade é investigar as consequências de se utilizar apenas *alguns* deles para definir $\mathbf{B}\dot{-}_{\alpha}$. Sendo mais objetivo, podemos tomar uma função de seleção *s* para escolher algum subconjunto não vazio de $\mathbf{B}\bot\alpha$. Tal processo é chamado de "contração de intersecção parcial" (do inglês *partial meet contraction*) e é definida da seguinte maneira:

$$\mathbf{B}\dot{-}_{\alpha} = \bigcap s(\mathbf{B}\bot\alpha)$$

A ideia por trás da contração de intersecção parcial é que a função *s* selecione o "melhor candidato" em $\mathbf{B}\bot\alpha$ para definir $\mathbf{B}\dot{-}_{\alpha}$. Tal função de seleção para os "melhores candidatos" de $\mathbf{B}\bot\alpha$ é um dos aspectos mais relevantes da teoria AGM, e voltaremos a ela no final desta seção. Mas antes de especificar o funcionamento da função *s*, vamos assumir que esse é o modelo correto para as operações de contração e mostrar como a "contração de intersecção parcial" corresponde a um conjunto de postulados (axiomas) para a teoria AGM, e como podemos derivar a noção de revisão a partir disso.

4.2.1.1 A sintaxe

No artigo canônico da teoria AGM, Alchourrón *et al.* (1985), encontramos um teorema da representação garantindo que a contração de intersecção parcial satisfaz seis axiomas para contração que podem ser formulados independentemente (isto é, possuem uma justificação própria) do método de intersecção parcial. Os postulados são os seguintes (GÄRDENFORS, 1988, p. 61-63):

($\mathbf{B}^{\div}1$)	$\forall\alpha \in \mathcal{L}, \mathbf{B}^{\div}_{\alpha}$ é um conjunto de crenças;	(*Fechamento*)
($\mathbf{B}^{\div}2$)	$\mathbf{B}^{\div}_{\alpha} \subseteq \mathbf{B}$;	(*Sucesso*)
($\mathbf{B}^{\div}3$)	Se $\alpha \notin \mathbf{B}$, então $\mathbf{B}^{\div}_{\alpha} = \mathbf{B}$;	(*Inclusão*)
($\mathbf{B}^{\div}4$)	Se não é o caso que $\vdash \alpha$, então $\alpha \notin \mathbf{B}^{\div}_{\alpha}$;	(*Preservação*)
($\mathbf{B}^{\div}5$)	Se $\alpha \in \mathbf{B}$, então $\mathbf{B} \subseteq (\mathbf{B}^{\div}_{\alpha})^{+}_{\alpha}$;	(*Consistência*)
($\mathbf{B}^{\div}6$)	Se $\vdash \alpha \leftrightarrow \beta$, então $\mathbf{B}^{\div}_{\alpha} = \mathbf{B}^{\div}_{\beta}$.	(*Extensionalidade*)

Tais postulados são reconhecidos como os axiomas de contração para uma teoria de revisão de crença AGM. Ou nas palavras de Gärdenfors (1988, p. 61), postulados de racionalidade baseados no critério de economia informacional.

Resumidamente, o primeiro postulado garante que podemos gerar novos conjuntos de crença contraindo qualquer sentença; o segundo exige que nenhuma informação nova seja adicionada, apenas retirada; o terceiro garante que, se α não pertence ao conjunto de crenças, remover α não produz qualquer efeito; o quarto garante que a sentença removida não deve pertencer mais ao conjunto de crenças, a menos que ela seja uma verdade lógica; o quinto postulado exige que todas crenças em **B** sejam recuperadas depois de primeiramente contraídas e em seguida adicionadas no que diz respeito a mesma informação; e por último, o sexto postulado afirma que proposições logicamente equivalentes resultarão em conjuntos contraídos idênticos.

Como vimos no início do presente tópico, um processo de revisão de crença sempre envolve um tipo de contração. Assim, uma revisão na teoria AGM é inspirada na simples observação de que, se p não pode ser consistentemente adicionada em **B**, é porque **B** contém $\neg p$ dentre suas crenças. Desse modo, para revisar **B** após receber p como evidência, primeiramente precisamos remover $\neg p$. Isso quer dizer que podemos definir uma função de revisão $*$ a partir dos processos de contração e expansão. A seguinte igualdade é conhecida por *Identidade de Levi*:

$$\mathbf{B}^{*}_{\alpha} = (\mathbf{B}^{\div}_{\neg\alpha})^{+}_{\alpha}$$

Uma vez que é possível construir funções de revisão a partir de funções de contração mais expansão, podemos agora apresentar o conjunto de postulados para revisão (que possuem uma forte semelhança com os axiomas de contração no que diz respeito ao papel de cada um deles):

($\mathbf{B}^*1$)	$\forall\alpha \in \mathcal{L}, \mathbf{B}^*_\alpha$ é um conjunto de crenças;	(*Fechamento*)
($\mathbf{B}^*2$)	$\alpha \in \mathbf{B}^*_\alpha$;	(*Sucesso*)
($\mathbf{B}^*3$)	$\mathbf{B}^*_\alpha \subseteq \mathbf{B}^+_\alpha$;	(*Inclusão*)
($\mathbf{B}^*4$)	Se $\neg\alpha \notin \mathbf{B}$, então $\mathbf{B}^+_\alpha \subseteq \mathbf{B}^*_\alpha$;	(*Preservação*)
($\mathbf{B}^*5$)	$\mathbf{B}^*_\alpha = \mathbf{B}_\perp$ *sse* $\vdash \neg\alpha$;	(*Consistência*)
($\mathbf{B}^*6$)	Se $\vdash \alpha \leftrightarrow \beta$, então $\mathbf{B}^*_\alpha = \mathbf{B}^*_\beta$.	(*Extensionalidade*)

Assim como os postulados para contração, os postulados para revisão também podem ser fundamentados por si só baseados no critério de economia informacional e nos requisitos racionais (sincrônicos) para crença apresentados anteriormente. Do mesmo modo, agora é possível fazer o caminho inverso e definir uma função de contração a partir da revisão. Tal resultado ficou conhecido como *Identidade de Harper*:

$$\mathbf{B}^{\div}_\alpha = \mathbf{B} \cap (\mathbf{B}^*_{\neg\alpha})$$

Com isso, temos um conjunto de postulados para os três processos fundamentais para mudança de visão: expansão, contração e revisão. Tais postulados formam o que conhecemos como uma teoria AGM de revisão de crença (em sua versão elementar). Ao mesmo tempo, é possível chegar à mesma teoria a partir da estratégia de *contração de intersecção parcial*. Mas ainda existem mais dois postulados, tanto para contração quanto para revisão, para termos uma teoria AGM completa. Mas para apresentar os dois postulados restantes é importante especificar seus elementos correspondentes na construção (não-axiomática) de uma função de contração de intersecção parcial.

4.2.1.2 De volta à *Contração de Intersecção Parcial*

Como prometido, vamos agora especificar o funcionamento da função de seleção *s* na contração de intersecção parcial. Sem tal especificação, somos incapazes de obter uma unicidade de elementos geradores do conjunto $\mathbf{B}^{\div}_\alpha$. Consequentemente, sem uma definição precisa do processo de contração somos incapazes de especificar uma estratégia de revisão.

Como dissemos anteriormente, a contração de intersecção parcial seleciona em $\mathbf{B}\perp\alpha$ o melhor candidato a $\mathbf{B}^{\div}_\alpha$. Mas e se houver mais de um candidato a melhor subconjunto? A proposta então é que o agente intersecccione todos os melhores candidatos para obter um conjunto $\mathbf{B}^{\div}_\alpha$ ainda menor. Contudo, ainda precisamos entender o que significa ser o "melhor candidato" selecionado pela função *s*. Para Gärdenfors (1988, p. 80), a função *s* deve escolher o subconjunto em $\mathbf{B}\perp\alpha$ que seja "mais epistemicamente entrincheirado" (do inglês, *epistemologically most entrenched*), isto é, os candidatos de maior valor epistêmico. Assim, uma sentença pertence a $\mathbf{B}^{\div}_{\alpha}$ se, e somente se, ela pertence a todos os subconjuntos maximais de $\mathbf{B}\perp\alpha$ que são mais epistemicamente entrincheirados. Tal noção de "entrincheiramento" é simplesmente uma forma de rotular crenças através de seu valor epistêmico. E isso é considerado um dos pontos mais importantes em uma teoria AGM. Nas palavras de Gärdenfors (1988, p. 88), "tomo a noção de entrincheiramento epistêmico como sendo mais fundamental do que a noção de função de contração ou revisão".

Visando determinar quais são os subconjuntos que serão selecionados pela função *s*, precisamos assumir algum *ranqueamento* em $\mathbf{B}\perp\alpha$, ou seja, uma ordem dos elementos mais epistemicamente entrincheirados. Para isso, é comum se assumir uma relação binária $\preccurlyeq$ sobre sentenças de tal modo que, $\alpha \preccurlyeq \beta$ significa que α é ao menos tão bom quanto β, no que diz respeito a ordem $\preccurlyeq$ de ranqueamento epistêmico. Assim, podemos definir a função *s* da seguinte maneira:

DEFINIÇÃO 24 (Função de Seleção)**.**

$$s(\mathbf{B}\perp\alpha) := \{\mathbf{B}' \in (\mathbf{B}\perp\alpha) : \forall\, \mathbf{B}'' \in (\mathbf{B}\perp\alpha), \mathbf{B}' \preccurlyeq \mathbf{B}''\}$$

É importante notar que, até agora, não foi especificada nenhuma propriedade da relação $\preccurlyeq$, apenas que seja uma relação binária. Segundo Gärdenfors (1988, p. 82), um requisito mínimo é que $\preccurlyeq$ seja *transitiva*. Além disso, se desejamos que $\preccurlyeq$ seja uma relação de ordem completa, e não apenas parcial, precisamos exigir que ela seja uma relação *total*. Isso significa que, para todo subconjunto $\mathbf{B}'$ e $\mathbf{B}'' \in (\mathbf{B}\perp\alpha)$, temos que ou $\mathbf{B}' \preccurlyeq \mathbf{B}''$ ou $\mathbf{B}'' \preccurlyeq \mathbf{B}'$. Com isso, qualquer operação de contração determinada por uma intersecção parcial, com uma função *s* que escolhe as opções mais bem ranqueadas por uma relação binária transitiva e total, é chamada de *contração de intersecção parcial transitivamente relacional e total* (GÄRDENFORS, 1988, p. 82).

A partir da noção completa de um procedimento de contração (com um ranqueamento de sentenças mais epistemicamente entrincheiradas), podemos completar o conjunto de postulados para contração na teoria AGM. Gärdenfors (1988, p. 81-82) mostrou que tal procedimento para contração satisfaz os seis postulados já apresentados acima, e mais os dois seguintes (também conhecidos como *Postulados suplementares de Gärdenfors*):

$$(\mathbf{B}^{\div}7) \qquad \mathbf{B}^{\div}_{\alpha} \cap \mathbf{B}^{\div}_{\beta} \subseteq \mathbf{B}^{\div}_{\alpha\wedge\beta}$$

$$(\mathbf{B}^{\div}8) \qquad \text{Se } \alpha \notin \mathbf{B}^{\div}_{\alpha\wedge\beta}, \text{ então } \mathbf{B}^{\div}_{\alpha\wedge\beta} \subseteq \mathbf{B}^{\div}_{\alpha}$$

Traduzindo o que tais postulados querem dizer, ($\mathbf{B}^{\div}$ 7) garante que crenças que estão tanto em $\mathbf{B}^{\div}_{\alpha}$ quanto em $\mathbf{B}^{\div}_{\alpha}$, estão também em $\mathbf{B}^{\div}_{\alpha\wedge\beta}$; e ($\mathbf{B}^{\div}$ 8) afirma que a mudança mínima em **B** necessária para abandonar a conjunção $\alpha\wedge\beta$ está intimamente relacionada com a mudança mínima necessária para rejeitar o próprio α. Além disso, o último postulado irá assegurar que se α é rejeitado em $\mathbf{B}^{\div}_{\alpha\wedge\beta}$, isso significa que α é menos epistemicamente entrincheirado que β.

Novamente, de maneira semelhante, temos as versões correspondentes dos postulados de revisão:

$$(\mathbf{B}^*7) \qquad \mathbf{B}^*_{\alpha\wedge\beta} \subseteq (\mathbf{B}^*_\alpha)^+_\beta$$
$$(\mathbf{B}^*8) \qquad \text{Se } \neg\beta \notin \mathbf{B}^*_\alpha, \text{ então } (\mathbf{B}^*_\alpha)^+_\beta \subseteq \mathbf{B}^*_{\alpha\wedge\beta}$$

No tópico 4.3 veremos argumentos em defesa e contra tais postulados, e como eles se relacionam com o debate entre concepções distintas de PP diacrônico.

4.2.1.3 A semântica: O Sistema de Esferas de Grove

Uma estratégia alternativa de construção para revisão[9] de crença foi apresentado por Adam Grove (1988) a partir de um modelo de revisão com mundos possíveis, também chamado de *Sistema de Esferas*, que será apresentado resumidamente a seguir. Antes de apresentar a construção de Grove, é importante retomar alguns conceitos vistos anteriormente.

No primeiro capítulo dissemos que uma investigação sobre representações formais de crença pode tomar o conteúdo da crença de pelo menos dois modos: como sentenças ou como proposições. Desde então, trabalhamos com proposições como sendo o conteúdo da crença, e proposições são entendidas como conjuntos de mundos possíveis. Mundos possíveis são descrições completas, ou modos maximamente especificados, de como o mundo pode ser. Assim como faz Grove (1988), tomaremos aqui mundos possíveis como conjuntos maximais consistentes de $\mathcal{L}$, isto é, conjunto de sentenças que são consistentes e fechados logicamente, mas que deixam de satisfazer uma das duas condições assim que uma nova sentença é adicionada (o que acontece com as operações de Expansão e Revisão). Ainda complementando, mundos possíveis são versões menos "refinadas" de sentenças, uma vez que existem infinitos modos de representar o mesmo mundo possível por meio de sentenças em $\mathcal{L}$. Neste tópico abordaremos a teoria AGM de um ponto de vista proposicional, a partir da suposição de que proposições são conjuntos de mundos possíveis que podem ser descritos em $\mathcal{L}$. Portanto, o conjunto universo $W_\mathcal{L}$ é o conjunto de mundos em que interpretamos as sentenças em $\mathcal{L}$ de tal modo que, $w_i \models p$ se, e somente se, $p \in w_i$, para qualquer $w_i \in W_\mathcal{L}$.

Uma vez que relacionaremos sentenças e proposições, vamos precisar introduzir alguns elementos em uma nova notação. Sendo assim, para qualquer sentença

[9] Por meio da *Identidade de Harper* tal estratégia também pode ser ampliada para contrações.

α, denotaremos por $[\alpha]$ o conjunto de mundos possíveis onde α é uma sentença verdadeira. Ainda, para qualquer conjunto Γ de sentenças, denotaremos $[\Gamma]$ o conjunto de mundos possíveis que tornam as sentenças em Γ verdadeiras (por exemplo, $[\mathbf{P}]$ é a proposição expressa pelo conjunto particular de sentenças **P**). Mais precisamente, para uma sentença arbitrária α, temos que:

$$[\alpha] := \{w \in W_{\mathcal{L}} : w \models \alpha\}.$$

Previsivelmente, $[\mathbf{B}]$ é o conjunto de mundos possíveis onde todas as sentenças do conjunto de crenças **B** são verdadeiras.

Com isso, temos que toda proposição pode ser expressa por um conjunto de sentenças em $\mathcal{L}$, e *vice-versa*: Para cada proposição (conjunto de mundos) $A \in W_{\mathcal{L}}$, existe um conjunto arbitrário de sentenças $\Gamma \subseteq \mathcal{L}$ onde $A = [\Gamma]$ e $\Gamma = \bigcap A$, onde $\bigcap A$ é a intersecção de todo $w_i \in A$. Como um mundo possível contém todas as sentenças verdadeiras naquele mundo, dizemos que $\bigcap A$ seleciona o conjunto de sentenças Γ verdadeiras em w_i, tal que, para todo $w_i \in A$, temos que $w_i \models \Gamma$. Como consequência, para cada proposição $A \in W_{\mathcal{L}}$, temos que $A = [\bigcap A]$, e para cada conjunto de sentenças $\Gamma \subseteq \mathcal{L}$, temos que $\Gamma = \bigcap[\Gamma]$. Ainda, isso quer dizer que para cada sentença atômica $\alpha \in \mathcal{L}$, existe uma proposição (conjunto de mundos) $[\alpha]$ tal que $\alpha = \bigcap[\alpha]$[10]. Em outras palavras, uma sentença é intersecção dos mundos onde as respectivas sentenças são verdadeiras[11].

Podemos explorar agora alguns dos possíveis candidatos a conjunto de crenças **B**. Não surpreendentemente, o menor deles é o conjunto que contém apenas verdades lógicas. Vamos chamá-lo de $\mathbf{B}_0$. Pelas propriedades lógicas de um conjunto de crença, podemos construir extensões maximais de $\mathbf{B}_0$ ao adicionar novas sentenças[12]. Consequentemente, ao adicionar uma sentença p em $\mathbf{B}_0$ obtemos $Cn(\mathbf{B}_0 \cup \{p\})$, que é precisamente a definição de *expansão* na teoria AGM. Assim, dado um conjunto qualquer de crenças **B**, cada $w_i \in W_{\mathcal{L}}$ é uma extensão maximal consistente de $\mathbf{B}_0$. Vale ressaltar que o uso de mundos possíveis também é uma forma de descrever o quanto o agente acredita: quanto mais ele aprende (quanto mais sentenças são adicionadas

[10] Segundo Grove (1988, p. 158), isso é o caso sempre que a lógica subjacente for *compacta*, o que é o caso na lógica clássica.

[11] Para o leitor com certa dificuldade na compreensão dos passos dados no parágrafo, pode ser útil retomar no Exemplo 1 do capítulo anterior, e distribuir um conjunto de n sentenças em uma tabela com 2^n possibilidades, onde cada possibilidade representa um mundo possível. A partir da construção no exemplo fica mais fácil entender como $p = \bigcap[p]$.

[12] Relembrando, **B** possui todas as propriedades de uma *Teoria* em Lógica. Isso significa que existe a menor teoria, podemos chamar de T_0 que consiste apenas no conjunto de axiomas lógicos. Assim sendo, podemos chamar de **M** o conjunto de todas extensões maximais de T_0. Nesse caso, **M** é equivalente à $\mathcal{A}$, ou seja, é o conjunto de todas proposições fechadas sob operações de conjuntos. Em Grove (1988), todo trabalho é feito em termos de Teorias Lógicas, ao invés de conjuntos de crenças e mundos possíveis como fazemos aqui. Mas como o próprio autor reconhece, os resultados podem ser expressos de ambos os modos (GROVE, 1988, p. 160).

em **B**), menor é o número de mundos possíveis compatível com o que ele acredita. Logo isso ficará mais claro com as Figuras 9 e 10.

Uma maneira conveniente de interpretar tal relação entre conjuntos de crença e mundos possíveis é por meio de diagramas. Na Figura 9, qualquer ponto na superfície do retângulo representa um mundo possível (assumindo apenas um número finito de pontos), e a área em cinza representa o novo conjunto de crenças após a expansão, precisamente $[\mathbf{B}] \cap [p]$. Isso nos auxilia na interpretação de um procedimento de mudança de visão a partir de mundos possíveis. No caso mais simples, quando a nova sentença adquirida p é compatível com o conjunto inicial de crenças **B**, como vimos com a teoria AGM, uma expansão resulta em $Cn(\mathbf{B} \cup \{p\})$. Em termos de mundos possíveis, isso significa que $[\mathbf{B}]$ e $[p]$ possuem uma intersecção não-vazia, isto é, existe ao menos um mundo onde ambas são verdadeiras. Nesse caso, apenas abandonamos os elementos de $[\mathbf{B}]$ que são incompatíveis com p, ficando assim com um subconjunto de $[\mathbf{B}]$. E combinando os postulados da teoria AGM com operações sobre conjuntos de mundos, nesse caso temos que $\mathbf{B}^*_p = \mathbf{B}^+_p = \bigcap([\mathbf{B}] \cap [p])$. Informalmente, isso quer dizer que quando uma nova informação p é adquirida em uma expansão, o agente se torna mais opinado, e o número de mundos possíveis compatíveis com o que ele acredita diminui, uma vez que agora os mundos onde $\neg p$ é o caso devem ser excluídos do conjunto de crenças **B**.

Aqui se faz necessário chamar a atenção para a relação entre os diagramas (Figuras 9 e 10) representando mundos possíveis e conjunto de crenças **B**: uma maior área no retângulo representa um menor conjunto de crenças, uma vez que um pequeno conjunto de crenças é compatível com um maior número de mundos. Assim, o menor conjunto de crenças $\mathbf{B}_0 = Cn(\emptyset)$, isto é, o conjunto de verdades lógicas, é o menor conjunto subconjunto de todos os mundos possíveis. Por outro lado, o maior conjunto de crenças é $\mathbf{B}_\perp = Cn(\perp)$, pois contém todas as sentenças em $\mathcal{L}$, sendo assim o conjunto contraditório, que não é subconjunto de nenhum mundo possível, e não pode ser representado no diagrama (dado que um mundo possível é uma extensão maximal *consistente* de um conjunto de sentenças). Por fim, a maior extensão consistente maximal expressa em $\mathcal{L}$ é o conjunto $W_\mathcal{L}$, e consiste em toda área do retângulo do diagrama, ou seja, o conjunto consistente que não deixa nenhum mundo possível de fora.

Contudo, como vimos no tópico anterior, os casos mais interessantes de mudança de visão são aqueles que implicam uma contração do conjunto inicial de tal modo que satisfaça o requisito de consistência. Em outras palavras, quando $[\mathbf{B}]$ e $[\alpha]$ possuem uma intersecção vazia, isso significa que o resultado de $\mathbf{B}^*_\alpha$ deve se encontrar fora de $[\mathbf{B}]$, mas ainda assim deve ser um subconjunto de $[\alpha]$, uma vez que, pelo axioma ($\mathbf{B}^*2$) da teoria AGM, $\alpha \in \mathbf{B}^*_\alpha$. A ideia aqui é tornar no novo conjunto revisado o mais próximo possível de $[\mathbf{B}]$ (uma extensão de **B**), mas distante o suficiente para que

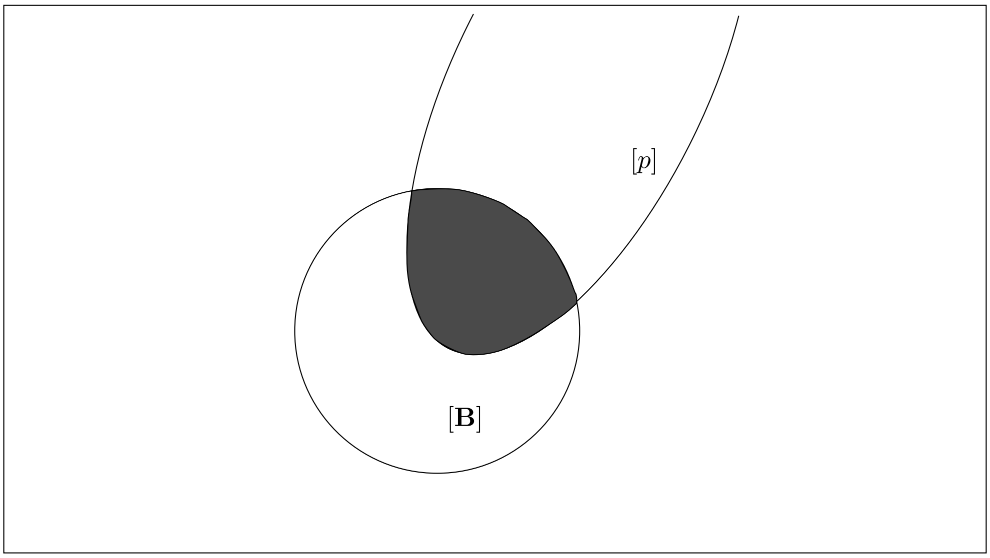

Figura 9 – Expansão de **B** por *p* em termos de mundos possíveis, ou seja, $\mathbf{B}_p^+$.

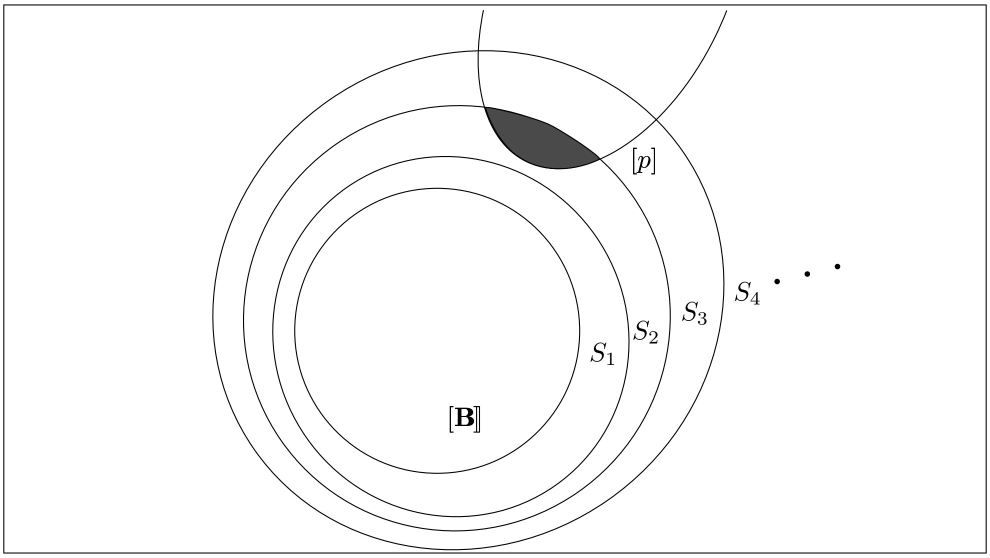

Figura 10 – Revisão de **B** por *p* em um sistema de esferas, ou seja, $\mathbf{B}_p^*$.

ele interseccione $[\alpha]$. Isso pode ser feito por meio de uma *coleção de superconjuntos* de [**B**]. Em um diagrama, isso pode ser visualizado como uma coleção de "esferas" em torno de [**B**], como podemos ver na Figura 10 (a área em cinza representa o novo conjunto de crenças após a revisão; note que o novo conjunto agora não pertence mais ao conjunto [**B**] inicial, uma vez que $[\mathbf{B}] \cap [p] = \emptyset$).

Vamos explorar um pouco mais o caso de revisão. Recordemos as possíveis extensões de **B**. Em termos lógicos, como já foi dito, o menor conjunto $\mathbf{B}_0$ pode ser entendido como o conjunto de axiomas lógicos de uma teoria. Assim como na lógica podemos adicionar novos axiomas não-lógicos e gerar novas Teorias (uma extensão dos axiomas), também podemos adicionar aqui novas sentenças e gerar um novo conjunto de crenças (uma extensão de $\mathbf{B}_0$). Conforme adicionamos mais sentenças e não removemos nenhuma, nosso conjunto se torna cada vez mais robusto, isso é, reduzindo cada vez mais o conjunto de mundos possíveis que tornam todas sentenças em **B** verdadeiras, como vimos no primeiro exemplo com uma expansão. Quando a nova informação é incompatível com **B**, somos obrigados a abrir mão das sentenças em **B** que são incompatíveis com a nova informação, e nos manter o mais próximo possível de **B**, pelo critério de economia informacional. Em termos proposicionais, precisamos ampliar o mínimo necessário o número de mundos em [**B**] até encontrar algum que torne a nova informação adquirida verdadeira. Podemos observar isso didaticamente por meio da Figura 10.

Como mencionamos no início do tópico, tal modelo de revisão de crença foi desenvolvido por Grove (1988) e é chamado de *Sistema de Esferas*. Em seu principal artigo sobre o tema, o autor demonstrou que o protocolo de revisão apresentado (informalmente) acima baseado em sistemas de esferas corresponde exatamente a revisão feita por meio de uma *contração de intersecção parcial transitivamente relacional e total* da teoria AGM, e satisfaz os axiomas ($\mathbf{B}^*1 - \mathbf{B}^*8$).

De acordo com Grove (1988, p. 160), um *Sistema de Esferas* é uma ordem de conjuntos sobre mundos possíveis[13]. Mais precisamente, um sistema de esferas é uma coleção $\mathbb{S}$ de *proposições* S_i subconjuntos de $W_{\mathcal{L}}$; formalmente, temos que $S \subseteq W_{\mathcal{L}}$ é uma proposição e $\mathbb{S}$ uma coleção de proposições no conjunto potência $\mathscr{P}(W_{\mathcal{L}})$. Cada mundo possível – extensão maximal de $\mathbf{B}_0$ – corresponde a uma situação que pode ser o caso no mundo real, no sentido que concorda com o mundo real onde as sentenças em **B** são verdadeiras. Um conjunto de crenças **B** corresponde ao quanto acreditamos saber sobre o mundo real, mais apropriadamente, que o mundo real é um daqueles contidos em [**B**]. Assim, um *Sistema de Esferas centrado em* [**B**] é uma coleção $\mathbb{S}$ de proposições em $W_{\mathcal{L}}$ que satisfaz as seguinte condições (GROVE, 1988, p. 159):

($\mathbb{S}$1) $\mathbb{S}$ é totalmente ordenado por $\subseteq$; ou seja, se S_1 e S_2 são elementos de $\mathbb{S}$, então $S_1 \subseteq S_2$ ou $S_2 \subseteq S_1$.

($\mathbb{S}$2) [**B**] é o $\subseteq$ –mínimo de $\mathbb{S}$; ou seja, dado $[\mathbf{B}] \in \mathbb{S}$, se $S_1 \in \mathbb{S}$ então $[\mathbf{B}] \subseteq S_1$.

($\mathbb{S}$3) $W_{\mathcal{L}} \in \mathbb{S}$, e é o maior elemento de $\mathbb{S}$.

($\mathbb{S}$4) Seja $\alpha \in \mathcal{L}$ uma sentença e S uma esfera qualquer em $\mathbb{S}$ que intersecciona $[\alpha]$, então existe uma menor esfera em $\mathbb{S}$ que intersecciona $[\alpha]$; ou seja, existe uma menor esfera S_i tal que $S_i \cap [\alpha] \neq \emptyset$, e para todo $S_j \in \mathbb{S}$, se $S_j \cap [\alpha] \neq \emptyset$ então $S_i \subseteq S_j$.

Desenvolvendo um pouco mais cada uma das condições acima, a primeira garante que o sistema de esferas é *aninhado*. O postulado ($\mathbb{S}$2) afirma que o sistema

[13] Semelhante à semântica de esferas para contrafactuais em Lewis (1973).

de esferas $\mathbb{S}$ é centrado em [**B**]. A terceira condição garante a existência do conjunto tautológico $W_{\mathcal{L}}$. E por fim, a última condição, em combinação com a terceira, diz que, se p é uma sentença consistente, então existe uma menor esfera S_p que intersecciona $[p]$.

Agora, para ser mais preciso na construção de um modelo de revisão AGM em termos de um Sistema de Esferas, precisamos de uma *função de seleção* que pegue os mundos compatíveis com uma determinada sentença p "mais próximos" de **B**, onde "proximidade" é entendido em termos do sistema de esferas $\mathbb{S}$. Desse modo, para cada sistema de esferas $\mathbb{S}$, associamos uma função c que seleciona a menor esfera dentre aquelas onde p é verdadeira. Mais precisamente, $c(\alpha) := [\alpha] \cap S_{\alpha}$, onde S_{α} é a menor esfera que intersecciona $[\alpha]$. Se $[\alpha]$ não intersecciona nenhuma esfera - o que ocorre somente quando $[\alpha] = \emptyset$ - então $c(\alpha)$ será o próprio $W_{\mathcal{L}}$. Seguindo o exemplo da Figura 10, seja $S_2 \in \mathbb{S}$ a menor esfera que intersecciona uma determinada sentença $[p]$. Assim definimos $c(p)$ como $[p] \cap S_2$, onde S_2 é a primeira esfera (fora de [**B**]) que intersecciona a proposição expressa por p.

Assim, dado um conjunto de crenças **B** e um sistema de esferas $\mathbb{S}$ centrado em [**B**], podemos definir uma função de revisão do seguinte modo:

$$\mathbf{B}^*_{\alpha} = \bigcap c(\alpha).$$

E esse é precisamente o resultado mais relevante do artigo de Grove: um teorema da representação sobre a correspondência entre um sistema de esferas e os axiomas de revisão na teoria AGM. Para isso, ele demonstrou o seguinte par de teoremas (GROVE, 1988, p. 161):

TEOREMA 4.2.1. *Seja $\mathbb{S}$ um sistema de esferas centrado em $[\boldsymbol{B}]$, para um conjunto de crenças **B**. Se, para cada sentença $\alpha \in \mathcal{L}$, $\boldsymbol{B}^*_{\alpha}$ é definido como $\bigcap c(\alpha)$, então o resultado da função de revisão no sistema de esferas irá satisfazer os axiomas $(\boldsymbol{B}^*1) - (\boldsymbol{B}^*8)$ da teoria AGM.*

TEOREMA 4.2.2. *Seja $* : \boldsymbol{B} \times \mathcal{L} \longrightarrow \boldsymbol{B}$ uma função qualquer que satisfaça $(\boldsymbol{B}^*1) - (\boldsymbol{B}^*8)$. Então, para qualquer conjunto (fixo) de crenças **B**, existe um sistema de esferas $\mathbb{S}$ que é centrado em $[\boldsymbol{B}]$ e que satisfaz $\boldsymbol{B}^*_{\alpha} = \bigcap c(\alpha)$ para qualquer $\alpha \in \mathcal{L}$.*

A demonstração desse resultado será importante para outros teoremas que veremos mais à frente, e uma versão do resultado original está reproduzida no Apêndice B para o leitor que queira compreender mais detalhadamente os dois teoremas.

Para finalizar, podemos também apresentar de modo equivalente uma semântica para a teoria AGM por meio de um ranqueamento total entre mundos possíveis, como mostrou Grove (1988). Isso é possível uma vez que, como vimos acima na propriedade $\mathbb{S}1$, um sistema de esferas $\mathbb{S}$ possui uma ordenação completa. Isso significa que o conjunto de crenças **B** é envolto de uma coleção de proposições ranqueadas que podem ser chamadas à participar como membros do conjunto caso uma revisão particular à exija. Assim, seja $\leq_{\mathbb{S}}$ uma relação de ordem entre mundos que satisfaz as seguintes propriedades (GROVE, 1988, p. 160):

($\leq_{\mathbb{S}}$ 1) $\leq_{\mathbb{S}}$ é conectado (ou totalmente ordenado); ou seja, para todo $w_i \in W_{\mathcal{L}}$, $w_1 \leq_{\mathbb{S}} w_2$, ou $w_2 \leq w_1$.

($\leq_{\mathbb{S}}$ 2) $\leq_{\mathbb{S}}$ é transitiva.

($\leq_{\mathbb{S}}$ 3) Se $[\alpha] \neq \emptyset$, então existe um $w_1 \in [\alpha]$ tal que $w_1 \leq_{\mathbb{S}} w_i$, para todo $w_i \in [p]$; ou seja, existe o "melhor" ($\leq_{\mathbb{S}}$ *–minimal*) entre tais mundos.

($\leq_{\mathbb{S}}$ 4) w_1 é $\leq_{\mathbb{S}}$ *–minimal* (ou seja, $w_1 \leq_{\mathbb{S}} w_i$; para todo $w_i \in W_{\mathcal{L}}$) se, e somente se, $w_1 \in [\mathbf{B}]$.

Esse sistema é relevante para revisão no seguinte sentido: a relação $\leq_{\mathbb{S}}$ deve ser interpretada como uma medida de quão "compatível" são os mundos alternativos em relação a nossas crenças atuais. Isso se assemelha a noção, vista anteriormente, de entrincheiramento epistêmico na teoria AGM. Como veremos, principalmente a primeira condição ($\leq_{\mathbb{S}}$ 1) nos será importante na sequência do capítulo quando falarmos sobre a principal alternativa à teria AGM para governar a mudança de visão. Mas antes disso, precisamos introduzir essa proposta alternativa à teoria AGM, e é o que faremos a seguir.

4.2.2 Uma Teoria do Raciocínio Não-Monotônico

Assumimos aqui, até então, certa familiaridade do leitor com a relação de consequência na lógica clássica, mas agora será importante retomar alguns pontos e avançar um pouco em alguns outros. Primeiramente, a lógica clássica utiliza uma linguagem formal da qual um conjunto (finito) qualquer Γ de sentenças (ou fórmulas bem formadas) é construído a partir de uma lista de letras elementares, do tipo p, q, r, etc., e conectivos binários do tipo $\wedge, \vee$ e conectivos unários como $\neg$. Tais conectivos podem ser entendidos em termos de suas tabelas de verdade, e o conjunto de todas as sentenças da linguagem é chamado de $\mathcal{L}$.

Para capturar a noção de consequência dedutiva, dizemos que $\alpha \vdash \beta$ sempre que β se *seguir logicamente* de α. Do mesmo modo, dizemos que $\Gamma \vdash \alpha$ sempre que α for uma consequência dedutiva do conjunto de sentenças $\gamma \in \Gamma$. Assim, uma consequência lógica é uma *relação* entre fórmulas. Mas também podemos interpretar consequência lógica como uma *operação* entre fórmulas, onde $Cn(\cdot)$ é uma operação de consequência dedutiva clássica e é definida como $Cn(\Gamma) = \{\alpha : \Gamma \vdash \alpha\}$.

Por fim, sabemos que a relação de consequência lógica clássica $\vdash$ satisfaz uma vasta lista de propriedades, algumas mais facilmente aceitas que outras. Uma que nos interessa particularmente é a propriedade de *Monotonicidade*: Se $\Gamma \vdash \alpha$ e $\Gamma \subseteq \Delta$, então $\Delta \vdash \alpha$. O estudo de *lógicas não-monotônicas* consiste basicamente no estudo de relações de consequência que não satisfazem *Monotonicidade*[14]. Mas isso ainda não diz muito sobre do que se trata exatamente uma lógica não-monotônica. Por isso, vamos dar um passo atrás e tentar compreender a motivação de fundo para tudo isso.

4.2.2.1 O início

A origem do estudo de lógicas não-monotônicas está ligada à insatisfação de alguns lógicos e cientistas da computação com as limitações da consequência lógica clássica para lidar com inferências a partir de premissas falíveis (ou anuláveis). O exemplo canônico[15] consiste no seguinte caso de inferência a partir de duas premissas: "Pássaros voam" e "Tweety é um pássaro"; disso concluímos que Tweety pode voar. Nesse caso, a premissa "pássaros voam" é interpretada como a informação "pássaros *normalmente* voam". Sendo assim, assumimos que Tweety pode voar na medida em que não temos nenhuma outra informação sobre Tweety capaz de mudar nossa opinião sobre a conclusão em questão. Mas, na medida em que uma nova informação, de que "Tweety é um pinguim", é adicionada ao conjunto de premissas, a conclusão anterior é revogada, mesmo que nenhuma das premissas tenha sido excluída (ou contraída). Esse é, por essência, um caso de raciocínio não-monotônico.

Recordemos que, na lógica clássica, se α é uma consequência lógica de β, ou seja, $\beta \vdash \alpha$, então α também é consequência de $\beta \wedge \gamma$, para uma sentença γ qualquer, isto é, $\beta \wedge \gamma \vdash \alpha$. E, como vimos, essa é precisamente a propriedade de monotonicidade. Desse modo, a motivação dos teóricos da lógica não-monotônica é a busca por uma relação de consequência que dê conta de casos atípicos, como é o caso de pássaros que não voam.

No exemplo acima, assumimos que, na ausência se informações contrárias, Tweety é um pássaro típico, e portanto, pode voar. É exatamente essa "ausência de informação contrária" que faz deste um caso de inferência não-monotônica. Isso também pode ser chamado de "raciocínio por ausência" (do inglês *reasoning by default*). O raciocínio por ausência torna, muitas vezes, um sistema de inferências lógicas mais simples, e está relacionado com o que Reiter (1987, p. 158) chama de "suposição de mundo fechado". Esta consiste na tese de que um sistema de raciocínio não-monotônico tacitamente supõe que a base de dados inicial é completa, ou seja, na ausência de uma informação específica sobre Tweety ser um pássaro "incomum", assumimos como falsa a afirmação de que "Tweety é um pinguim". Nesse caso, nossa

[14] Muitas vezes aqui utilizamos o termo "Lógica Clássica" em contraste a suas versões não-monotônicas, mas isso não quer dizer que estamos assumindo aqui que lógicas não-monotônicas são lógicas não-clássicas. Sobre o assunto, ver David Makinson (2003, p. 72).

[15] É assim que Reiter (1987) se refere a tal exemplo em um dos artigos mais relevantes para o início da literatura sobre raciocínio não-monotônico.

base de conhecimento contém a informação de que pinguins não voam, embora isso não apareça inicialmente nas premissas. Logo, uma vez que a informação de que Tweety é um pinguim é adquirida, não hesitamos em concluir que Tweety não pode voar, e isso ocorre porque a premissa "pássaros voam" é uma afirmação menos específica do que "pinguins não voam".

4.2.2.2 O raciocínio não-monotônico e o Sistema Preferencial

É importante aqui mencionar que o conceito de Lógica não-monotônica cobre uma enorme classe de teorias. Dentre as mais famosas, temos a Lógica *Default*, Circunscrição (Circumscription), Lógica Autoepistêmica, Redes de Sucessão (*Inheritance Network*), entre outras[16]. Nosso objetivo aqui será abordar a lógica não-monotônica para uma teoria de revisão de crença, como aparece fundamentalmente em David Makinson (2003), Shoham (1987) e Stalnaker (1994). A relevância de tal abordagem é significativa o suficiente para que, muita vezes, o estudo de lógicas não-monotônicas seja referido como simplesmente o estudo do *raciocínio* não-monotônico. E não é difícil notar a conexão entre relações não-monotônicas e raciocínio. No processo cotidiano de mudança de visão, na maioria das vezes temos informação incompleta sobre os eventos relevantes, e geralmente nos baseamos em suposições para dar prosseguimento ao raciocínio em questão. Como vimos, as relações de consequência clássicas $\vdash$ não conseguem retratar essa infinidade de casos de raciocínio com informação incompleta.

Seguindo esse entendimento, uma relação de consequência não-monotônica é uma função de revisão $\alpha \mid\!\sim \beta$, que significa que o agente está disposto a aceitar β ao receber α como evidência[17]. Mas, até o presente momento, apenas especificamos não-monotonicidade como aquilo que não é monotônico. Para ir além, precisamos especificar qual o conjunto de regras que uma lógica não-monotônica deve satisfazer. Em outras palavras, precisamos estabelecer as propriedades que caracterizam uma relação do tipo $\alpha \mid\!\sim \beta$.

Embora não exista uma reposta consensual sobre o desafio do parágrafo acima, em um artigo de 1990, o trio Sarit Kraus, Daniel Lehmann, e Menachem Magidor ficou reconhecido por sistematizar um conjunto geral de regras, e uma caracterização semântica para uma lógica não-monotônica. Tais regras ficaram conhecidas como *Propriedades KLM* (fazendo referência às iniciais dos autores). E tal sistematização nos proporciona uma investigação sobre a função de cada postulado para um objetivo particular de uma estrutura não-monotônica, uma vez que parece não fazer sentido falar em "A" lógica não-monotônica. Nas palavras de David Makinson (2003, p. 72):

---

[16] Para uma apresentação ampla dessas teorias ver Gabbay *et al.* (1994), respectivamente os capítulos 4, 6, 5 e 3.

[17] O mesmo pode ser feito com uma *operação* de consequência não-monotônica *Inf* definida como $Inf(\alpha) = \{\beta : \alpha \mid\!\sim \beta\}$, como faz Rott (2001, cap. 4).

> Dada a unicidade da inferência clássica, é natural para o estudante, intrigado ao ver os diversos tipos de consequência não-monotônica, se perguntar: Qual seria a verdadeira inferência não-monotônica? Qual é a relação de consequência não-monotônica correta? Qual é aquela que utilizamos na prática, mesmo que ainda estudemos outras? A resposta é que não existe. Não há uma única relação de consequência não-monotônica, mas uma quantidade indefinida delas.

Dito isso, a partir de Kraus *et al.* (1990), apresentaremos agora um menu de princípios, ou axiomas, para uma relação de consequência que pretenda ser chamada de não-monotônica. Um sistema não-monotônico que satisfaz o seguinte conjunto de propriedades em sua relação de inferência $\mid\!\sim$ é chamado de *Sistema Preferencial*, daqui em diante, **P**. As regras a seguir podem ser vistas como uma sintaxe de **P**:

(*KLM*1) $\alpha \mid\!\sim \alpha$	Reflexividade
(*KLM*2) Se $\vdash \alpha \leftrightarrow \beta$ e $\alpha \mid\!\sim \gamma$, então $\beta \mid\!\sim \gamma$	Equivalência Lógica à Esquerda
(*KLM*3) Se $\vdash \alpha \rightarrow \beta$ e $\gamma \mid\!\sim \alpha$, então $\gamma \mid\!\sim \beta$	Enfraquecimento à Direita
(*KLM*4) Se $\alpha \wedge \beta \mid\!\sim \gamma$ e $\alpha \mid\!\sim \beta$, então $\alpha \mid\!\sim \gamma$	Corte
(*KLM*5) Se $\alpha \mid\!\sim \beta$ e $\alpha \mid\!\sim \gamma$, então $\alpha \wedge \beta \mid\!\sim \gamma$	Monoticidade Cautelosa
(*KLM*6) Se $\alpha \mid\!\sim \gamma$ e $\beta \mid\!\sim \gamma$, então $\alpha \vee \beta \mid\!\sim \gamma$	Ou

É importante aqui uma breve explicação de cada uma das propriedades acima, baseado em Kraus *et al.* (1990, p. 176–8), interpretando a relação $\mid\!\sim$ como "se ..., normalmente ...". *Reflexividade* é a mais trivial das propriedades e parece ser satisfeita por qualquer raciocínio baseado em alguma noção de consequência lógica. *Equivalência Lógica à Esquerda* já apresenta uma influência da lógica subjacente em $\vdash$, e diz que sentenças logicamente equivalentes possuem exatamente as mesmas consequências. *Enfraquecimento à Direita* diz que deve-se aceitar como consequência "normal" (ou seja, $\mid\!\sim$) tudo o que é logicamente implicado pelo que é aceito como consequência "normal". *Corte* diz que se uma sentença na premissa é "normalmente" implicada por outra sentença na premissa, então a última já é suficiente para garantir a conclusão. Em outras palavras, *Corte* assegura que não há perda de poder inferencial ao longo de uma cadeia de derivações. *Monotonicidade Cautelosa* diz que, ao receber uma nova informação cuja verdade já poderia ser concluída "normalmente" pelas premissas, a nova informação não deve invalidar conclusões prévias.

As regras *KLM*1 – *KLM*5 podem ser tomadas como as propriedades fundamentais de qualquer sistema que queira ser considerado um sistema lógico. Para Kraus *et al.* (1990, p. 176), um sistema não-monotônico que satisfaz apenas *KLM*1 – *KLM*5 pode ser chamado de um sistema *Cumulativo*. Um sistema Cumulativo é mais fraco que um sistema Preferencial, o qual pode ser construído a partir do primeiro adicionando *KLM*6.

O conjunto de axiomas *KLM*1 – *KLM*6 constitui uma axiomatização do Sistema Preferencial **P**. Uma das propriedades interessantes, que merece um destaque especial, que conseguimos derivar ao adicionar (*KML*6) ao sistema Cumulativo, é o que Kraus *et al.* (1990) chamam de *Metade Difícil do Teorema da Dedução*, ou também conhecida como *Condicionalização*:

(*KLM*7) Se $\alpha \wedge \beta \mid\!\sim \gamma$, então $\alpha \mid\!\sim \beta \rightarrow \gamma$ Condicionalização

Para demonstrar tal resultado, basta aplicar (*KLM*3) na premissa considerando que $\vdash \gamma \rightarrow (\beta \rightarrow \gamma)$ e $\vdash \gamma \rightarrow (\neg\beta \rightarrow \gamma)$, chegando tanto em $\alpha \wedge \beta \mid\!\sim \beta \rightarrow \gamma$ como em $\alpha \wedge \neg\beta \mid\!\sim \beta \rightarrow \gamma$. Por (*KLM*6), temos que $(\alpha \wedge \beta) \vee (\alpha \wedge \neg\beta) \mid\!\sim \beta \rightarrow \gamma$, e uma vez que $(\alpha \wedge \beta) \vee (\alpha \wedge \neg\beta)$ é logicamente equivalente à α, por (*KLM*2) podemos concluir que $\alpha \mid\!\sim \beta \rightarrow \gamma$.

Uma vez que **P** satisfaz tais regras, agora podemos buscar uma interpretação semântica para tal sistema. Essa interpretação é fundamentalmente baseada na ideia de *Modelos Preferenciais* de Shoham (1987, p. 72). Abaixo apresento resumidamente as principais características de uma semântica para **P**. O desenvolvimento a seguir é uma adaptação dos resultados originais, os quais envolvem um nível de profundidade técnica que foge aos nossos propósitos aqui. Os resultados iniciais são encontrados em Shoham (1987), e ampliados posteriormente em Kraus *et al.* (1990).

Primeiramente, vamos retomar alguns conceitos clássicos de consequência *semântica*[18]. Seja v uma *valoração* sobre as sentenças em $\mathcal{L}$, definida como $v : \mathcal{L} \longrightarrow W_{\mathcal{L}}$. Para simplificar a notação, denotaremos $v(\alpha)$ como $[\alpha]$. Desse modo, como anteriormente, temos que $[\alpha]$ é o subconjunto de $W_{\mathcal{L}}$ onde α é verdadeira, ou seja, α é verdadeira em um mundo w_i se, e somente se, $w_i \in [\alpha]$. Do mesmo modo, dizemos que $[\Gamma]$ é o subconjunto de $W_{\mathcal{L}}$ onde toda sentença $\gamma \in \Gamma$ é verdadeira. Assim, podemos definir uma relação $\models$ de consequência semântica clássica como:

DEFINIÇÃO 25. $\Gamma \models \alpha$ *sse*, para toda valoração v, é o caso que $[\Gamma] \subseteq [\alpha]$.

Agora avançando um pouco mais em direção à semântica de um Sistema Preferencial, chamamos de uma *Estrutura Preferencial* $\mathcal{S}$ a tripla $\langle W_{\mathcal{L}}, v, \prec \rangle$, onde $W_{\mathcal{L}}$ é um conjunto de mundos possíveis descritos em $\mathcal{L}$, v uma valoração, e $\prec$ uma relação binária qualquer. Ainda, dizemos que α é verdadeira em w_i se, e somente se, $w_i \in [\alpha]$. Ainda, $\prec$ deve ser entendida como uma relação de preferência entre os elementos de $W_{\mathcal{L}}$, ou seja, entre mundos possíveis. Assim, dizemos que $w_i \prec w_j$ quando o mundo w_i é preferido sobre o mundo w_j. Agora podemos definir uma *relação de consequência preferencial*:

[18] Trabalharemos aqui com uma semântica de mundos possíveis.

DEFINIÇÃO 26. Dizemos que $w_i \models_{\prec} \alpha$ (w_i preferencialmente satisfaz α) se, e somente se, $w_i \in [\alpha]$ e não existe outro mundo w_j tal que $w_j \in [\alpha]$ e $w_j \prec w_i$.

Nesse caso dizemos que w_i é um *mundo* $\alpha -$ *minimal*.

Um dos elementos mais importantes na semântica de **P** é a relação de preferência $\prec$. Essa relação representa uma preferência do agente entre mundos distintos. Mas quais propriedades $\prec$ deve satisfazer? Dado $W_{\mathcal{L}}$ um conjunto finito[19], $\prec$ é uma relação de ordem parcial estrita sobre $W_{\mathcal{L}}$, ou seja, é uma relação irreflexiva e transitiva. Relembrando, para todo $w_i \in W_{\mathcal{L}}$, uma relação é: transitiva quando, se $w_1 \prec w_2$ e $w_2 \prec w_3$, então $w_1 \prec w_3$; e irreflexiva quando não é o caso que $w_i \prec w_i$. Dizemos que uma relação de ordem é *total* se, e somente se, para todo $w_i, w_j \in W_{\mathcal{L}}$, com $i \neq j$, $w_i \prec w_j$ ou $w_j \prec w_i$. E dizemos que uma relação de ordem é *parcial* se ela não é total.

Com isso uma relação de consequência não-monotônica em um Sistema Preferencial **P**, definida a partir da relação $\mid\!\sim$, é definida semanticamente em uma estrutura preferencial $\mathcal{S}$ como:

DEFINIÇÃO 27. $\alpha \mid\!\sim_{\mathcal{S}} \beta$ *sse* β é verdadeira em todo *mundo* $\alpha -$ *minimal*.

Com isso, o seguinte teorema representa uma condição de *Correção* para **P**. Ou seja, todos os axiomas (*KLM*1 – *KLM*7) são satisfeitos em uma estrutura Preferencial $\mathcal{S}$ (KRAUS *et al.*, 1990, p. 193):

TEOREMA 4.2.3 (Correção em **P**). *Seja $\prec$ é uma ordem estrita parcial sobre $W_{\mathcal{L}}$ tal como definida em acima, então $\mid\!\sim_{\mathcal{S}}$ satisfaz as propriedades básicas de um Sistema Preferencial (KLM1 – KLM7).*

E, por outro lado, também conseguimos o seguinte resultado (KRAUS *et al.*, 1990, p. 195-6):

TEOREMA 4.2.4 (Completude em **P**). *Suponha que exista uma relação de consequência $\mid\!\sim$ qualquer, que satisfaça os postulados* (*KLM*1 – *KLM*7)*. Então existe um estrutura $\mathcal{S}$ onde $\alpha \mid\!\sim \beta$ sse β é verdadeira em todo mundo $\alpha -$ minimal.*

Os dois teoremas acima formam um Teorema da Representação para uma relação de consequência preferencial:

TEOREMA 4.2.5 (Teorema da Representação em **P**). *Uma relação de consequência lógica é uma relação preferencial sse é definida a partir de uma estrutura preferencial $\mathcal{S}$.*

[19] Caso $W_{\mathcal{L}}$ seja infinito, $\prec$ deve ser uma relação binária que satisfaz o que ficou conhecido como condição de "suavidade" (do inglês, *smoothness condition*) (KRAUS *et al.*, 1990, p. 181-183): seja $\prec$ uma relação binária sobre $W_{\mathcal{L}}$, se $W_{\mathcal{L}}$ possui um elemento mínimo e ele é único, dizemos que $W_{\mathcal{L}}$ é suave. Dizemos que um Modelo Preferencial $\mathcal{S} = \langle W_{\mathcal{L}}, v, \prec \rangle$ satisfaz a condição de suavidade *sse* para todo $\alpha \in \mathcal{L}$, o modelo de α é suave, ou seja, o subconjunto de $W_{\mathcal{L}}$, onde $w_i \models \alpha$, é suave. A condição de suavidade é necessária para assegurar a validade de (*KLM*5), e se assemelha com a famosa "suposição de limite", presente em Lewis (1973).

A demonstração do teorema da representação decorre dos dois teoremas anteriores. Novamente, os resultados originais são apresentados em uma amplitude teórica que não acrescenta aos nossos propósitos específicos. Para o leitor interessado, encontramos a demonstração do resultado acima em Kraus *et al.* (1990, seção 5.3).

4.2.2.3 Uma versão mais forte de **P**

Para finalizar essa breve incursão não-monotônica, vamos apresentar aqui uma forma (controversa) de fortalecer o sistema **P**, a partir do seguinte postulado adicional:

($KLM8$) Se $\alpha \mid\!\sim \beta$ e $\alpha \not\mid\!\sim \neg\gamma$, então $\alpha \wedge \gamma \mid\!\sim \beta$ Monoticidade Racional

Essa regra é conhecida na literatura como *Monotonicidade Racional*. Esta expressa o fato de que ao adicionar uma nova informação no conjunto de crenças, alguma crença previamente adquirida só poderá ser removida se a nova informação gerar uma inconsistência com a crença prévia. Um sistema que satisfaz ($KLM1 - KLM8$) é chamado de um *Sistema Racional* **R**, e uma relação de consequência $\mid\!\sim$ que satisfaz ($KLM1$–$KLM8$) será uma *relação de consequência racional* $\mid\!\sim_{\mathbf{R}}$ (LEHMANN; MAGIDOR, 1992, sec. 3).

Resumidamente, uma semântica para um Sistema Racional **R** consiste em uma Estrutura $\mathcal{S}' = \langle W_{\mathcal{L}}, v, \prec_{\mathbf{R}} \rangle$, onde $\prec_{\mathbf{R}}$ é agora uma relação de ordem estrita e total, também chamada de relação de ordem *modular* (LEHMANN; MAGIDOR, 1992, p. 21). Diferentemente de um Sistema Preferencial, agora teremos um ordenamento completo, e não meramente parcial, no conjunto de mundos possíveis. Uma estrutura com tal propriedade também é chamada de um *ranqueamento*: cada mundo em $W_{\mathcal{L}}$ é ranqueado de tal forma que um mundo de menor ranqueamento é entendido como sendo um mundo "mais normal" do que um de maior ranqueamento. Em outras palavras, para cada mundo $w_i \in W_{\mathcal{L}}$, existe uma função bijetora entre $W_{\mathcal{L}}$ e um subconjunto finito dos números naturais, formando assim um ranqueamento completo. O restante segue-se igual ao modelo preferencial, e definimos uma relação de consequência não-monotônica para um Sistema Racional **R** em uma estrutura racional $\mathcal{S}'$ do seguinte modo:

DEFINIÇÃO 28. $\alpha \mid\!\sim_{\mathcal{S}'} \beta$ *sse* β é verdadeira em todo *mundo* $\alpha - minimal_{\boldsymbol{R}}$

Os seguintes resultados aparecem em Lehmann e Magidor (1992, seção 3.6):

TEOREMA 4.2.6 (Correção em **R**). *Se $\prec_{\boldsymbol{R}}$ uma ordem estrita total sobre $W_{\mathcal{L}}$, então $\mid\!\sim_{\mathcal{S}'}$ satisfaz as propriedades básicas de um Sistema Racional, isto é, Sistema Preferencial + Monotonicidade Racional ($KLM1 - KLM8$).*

TEOREMA 4.2.7 (Completude em **R**). *Suponha que exista uma relação de consequência* $\mid\!\sim$ *que satisfaz os postulados (KLM1 – KLM8). Então existe uma estrutura* $\mathcal{S}'$ *onde* $\alpha \mid\!\sim \beta$ *sse* β *é verdadeira em todo mundo* α *– minimal*$_{\boldsymbol{R}}$

Esses dois teoremas formam um Teorema da Representação para um Sistema Racional:

TEOREMA 4.2.8 (Teorema da Representação em **R**). *Uma relação de consequência lógica é uma relação Racional sse é definida a partir de uma estrutura preferencial* $\mathcal{S}'$.

Na próxima seção, mais precisamente Seção 4.3.3, veremos como a distinção entre um modelo preferencial e um modelo racional possui sérias implicações para a disputa sobre qual a melhor teoria formal para modelar nosso processo de mudança de visão qualitativa. Veremos com mais detalhes algumas razões pelas quais alguns autores tratam o axioma de Monotonicidade Racional como um princípio controverso.

4.2.3 Traduzindo uma pela outra

Finalmente, Para concluir essa imersão sobre uma lógica para a mudança de visão da crença, vamos apresentar brevemente algumas pontes que podemos construir entre as teorias vistas acima, e também sobre como podemos unificar a linguagem utilizada em acordo com o que viemos utilizando nos capítulos anteriores.

David Makinson e Gärdenfors (1991) apresentaram uma tradução entre operações de revisão de crença AGM e a relação de consequência não-monotônica $\mid\!\sim$. O seguinte princípio ficou conhecido como *Identidade de Makinson-Gärdenfors*, sendo α, β sentenças quaisquer:

$$\alpha \mid\!\sim_{\mathbf{B}} \beta \ \textit{sse} \ \beta \in \mathbf{B}^*_\alpha$$

Tal identidade representa a ideia de que, dada qualquer estratégia de revisão de crença $*$ em **B**, podemos definir uma relação de consequência não-monotônica $\mid\!\sim$ dependente de **B**, e vice-versa. Isso estabelece uma correspondência entre as duas abordagens.

O procedimento de tradução é formulado a partir dos seguintes passos (MAKINSON, D.; GÄRDENFORS, 1991, p. 189):

- **De** $\mid\!\sim$ **para** $*$: Simplesmente substitua cada parte elementar de $\alpha \mid\!\sim_{\mathbf{B}} \beta$ por $\beta \in \mathbf{B}^*_\alpha$.
- **De** $*$ **para** $\mid\!\sim$: Esse caso é um pouco mais delicado, uma vez que envolve a eliminação da variável **B**. Aqui são necessários três passos:
  (i) Observe se as condições sobre $*$ envolvem considerar a tradução de dois conjuntos de crenças distintos **B** e **B**$'$. Se sim, pare: tal condição não pode em geral ser traduzida. Se não envolve, siga para o próximo passo.

(ii) Observe se as condições sobre $*$ envolvem referências a teoria em ambas as formas revisadas e não revisadas de $\mathbf{B}^*_\alpha$ e **B**, respectivamente. Se envolve, então divida a condição em dois casos onde **B** é consistente e um caso limite onde **B** é inconsistente. No primeiro caso, substitua todas as ocorrências não-revisadas de **B** por $\mathbf{B}^*_\top$. No segundo caso, remova todas as ocorrências não-revisadas de **B** tratando as partes elementares $\alpha \in \mathbf{B}$ da condição a serem traduzidas como verdadeiras. Se não envolve, siga para o próximo passo.

(iii) Tome as saídas do passo (ii) e traduza-as substituindo todas partes elementares da forma $\beta \in \mathbf{B}^*_\alpha$ por $\alpha \mid\!\sim_{\mathbf{B}} \beta$.

Com essa ferramenta em mãos, seguindo David Makinson e Gärdenfors (1991, seção 4), podemos estabelecer algumas correlações (não-completas) entre o conjunto de postulado AGM ($\mathbf{B}^*1 - \mathbf{B}^*8$) e o conjunto de postulados (*KML*1 – *KLM*8). Deixando de lado alguns detalhes específicos, a seguinte lista de comparações pode ser apresentada[20]:

- ($\mathbf{B}^*1$) corresponde a ideia de que a relação não-monotônica é fechada sob operações lógicas.
- ($\mathbf{B}^*2$) corresponde ao postulado de *Reflexividade*.
- ($\mathbf{B}^*3$) é um caso particular da *Condicionalização*, embora não possua uma contraparte própria dentro do conjunto de postulados *KML*1 – *KLM*8.
- ($\mathbf{B}^*4$) é uma consequência da *Monotonicidade Racional*.
- ($\mathbf{B}^*5$) não é válida em nenhum dos sistemas não-monotônicos apresentados aqui.
- ($\mathbf{B}^*6$) corresponde à *Equivalência à Esquerda*.
- ($\mathbf{B}^*7$) corresponde à *Condicionalização*;
- ($\mathbf{B}^*8$) corresponde à *Monotonicidade Racional*.

Esse conjunto de relações entre as duas teorias será importante no tópico 4.3, onde voltaremos nossa atenção novamente para a busca de um Princípio Ponte que conecte crença (qualitativa) e probabilidade subjetiva, mas agora de um ponto de vista diacrônico.

Por fim, quero traçar alguns paralelos entre a linguagem que viemos utilizando até aqui para representar operações de crença $Bel(\cdot)$ e a linguagem que utilizamos nesse capítulo para apresentar a teoria AGM e os sistemas não-monotônicos. Isso

[20] Uma lista mais completa sobre tal correspondência pode ser encontrada em Rott (2001, p. 118).

nos ajudará a tornar mais simples e uniforme a linguagem utilizada na discussão que veremos no restante do capítulo.

Uma vez que *Bel*(A) significa que a proposição A pertence ao conjunto de crenças [**B**] do agente, dizemos que *Bel*($A|E$) representa uma operação sobre o conjunto de crenças ao receber uma nova informação E, em outras palavras, o agente possui uma *crença condicional* afirmando que está disposto a acreditar em A, se E for o caso. Em termos da teoria AGM, *Bel*($A|E$) := $A \in \mathbf{B}^*_E$, ou também, $\mathbf{B}^*_E \models A$. Por outro lado, um sistema não-monotônico com uma relação de consequência $\mid\!\sim$, toma *Bel*($A|E$) := $E \mid\!\sim_{\mathbf{B}} A$. Crenças condicionais receberão a devida atenção no início da próxima seção, e o uso das diferentes notações acima será intercalado, visando facilitar a compreensão.

Ainda, dizemos que *Bel*($A|E$) é uma operação com proposições, tal como definido no capítulo 1, Seção 1.2.1, onde o conjunto de crenças é determinado pelo conjunto de proposições acreditadas pelo agente. Como vimos anteriormente, podemos entender proposições em termos de sentenças, como o conjunto de mundos possíveis que torna exatamente a respectiva sentença verdadeira. Nesse sentido, o conjunto [**B**] de todas proposições acreditadas pelo agente é um conjunto de mundos possíveis compatíveis com todas a sentenças em **B**. Esse tipo de modelo para estados epistêmicos é chamado de “Modelo de Mundos Possíveis”. Relembrando, dado p uma sentença em $\mathcal{L}$, temos que $[p]$ representa o conteúdo da sentença p, ou seja, uma proposição, ou mais especificamente, o conjunto de mundos onde p é verdadeira.

Daqui em diante trataremos conjuntos de crenças como simplesmente conjuntos de proposições. Isso significa que os resultados da teoria AGM e dos sistemas não-monotônicos para um conjunto de sentenças acreditadas serão válidos para conjuntos de proposições acreditadas. Assim, por simplicidade de notação, daqui em diante, quando **B** aparecer em $\mathbf{B}^*_E$, ou em $E \mid\!\sim_{\mathbf{B}} A$, estarei tomando $\mathbf{B} = [\mathbf{B}]$, ou seja, o conjunto de proposições acreditadas pelo agente. Por fim, para simplificar e unificar a notação utilizada daqui em diante, denotaremos como *Bel*($\cdot \mid \cdot$) a crença em uma *proposição condicional*, de tal modo que *Bel*($A \mid E$) := $A \in \mathbf{B}^*_E = E \mid\!\sim_{\mathbf{B}} A$, sempre que $\mathbf{B}^*$ e $\mid\!\sim_{\mathbf{B}}$ cumprirem as exigências de tradução apresentadas acima. Tal simplificação na notação em operações com proposições condicionais facilitará nosso trabalho na discussão que veremos a seguir.

4.3 PRINCÍPIO PONTE COMO UM REQUISITO RACIONAL DINÂMICO

Relembrando, começamos este capítulo explorando a possibilidade de uma visão dinâmica da racionalidade e como requisitos racionais poderiam (e deveriam) ser aplicados diacronicamente em um contexto de mudança de visão. Uma vez que o princípio de condicionalização bayesiana é bem estabelecido como um requisito racional diacrônico para probabilidades subjetivas, a Seção 4.2 foi dedicada exclusivamente a requisitos racionais dinâmicos para a noção qualitativa de crença. O objetivo agora neste terceiro tópico será desenvolver e ampliar a investigação sobre um Princípio Ponte (PP) que conecte a racionalidade da crença e da probabilidade subjetiva, agora de um ponto de vista dinâmico. Isso quer dizer que não queremos que apenas nossas regras sincrônicas iniciais sejam satisfeitas por um PP apropriado, mas também as regras diacrônicas que foram apresentadas aqui, a saber, uma lógica para a mudança de visão de crença e o princípio de condicionalização bayesiana. Para isso, é importante ampliar nossa compreensão sobre o que há, e se há algo, em comum entre o princípio bayesiano da condicionalização e as duas teorias vistas acima sobre a lógica da crença.

Um leitor mais curioso já deve ter se questionado: qual a conexão teórica entre a condicionalização bayesiana e uma teoria lógica de revisão de crença, seja ela AGM ou não-monotônica? Em outras palavras, dado um conjunto [**B**] de crenças que contenha a proposição A, e dado uma proposição E qualquer, existe alguma interposição entre $Bel(A|E)$ e $P(A|E)$? Na medida em que todas elas possuem alguma semelhança com o que é expresso na linguagem natural em sentenças como "Se..., então...", uma maneira de interpretar tal relação seria assumir que elas modelam o mesmo tipo de fenômeno, a saber, os famosos "Condicionais". Devido à relevância do tema, dedicaremos a seguir alguns parágrafos para especificar um pouco melhor isso que vamos chamar de *crenças condicionais*.

É amplamente reconhecido na literatura sobre condicionais que o (assim chamado pelos lógicos) *condicional material* "$\rightarrow$" é bastante limitado para representar o "Se..., então..." da linguagem natural. Na lógica proposicional clássica, o significado de "$\rightarrow$" é determinado por sua tabela de verdade, e pode ser compreendido como expressando basicamente o seguinte: não é o caso que o antecedente é verdadeiro e o consequente é falso. Tal caracterização semântica do condicional material faz com que algumas consequências lógicas sejam aceitas como válidas, embora possuam uma profunda desconexão com nosso uso ordinário da linguagem. Por exemplo, se queremos entender o processo de raciocínio por traz da expressão "Se..., então...", o "$\rightarrow$" possui pouca, ou nenhuma, relevância quando o antecedente do condicional é falso, como em $\neg p \models p \rightarrow q$, o que faz com que qualquer sentença seja implicada materialmente por uma contradição; e também "$\rightarrow$" não possui muita utilidade quando o consequente é verdadeiro, como em $q \models p \rightarrow q$, uma vez que verdades lógicas são consequência de qualquer sentença.

Além da notável desconexão entre o uso do condicional material e o sentido do "Se..., então..." na linguagem natural, existem os famosos casos de *condicionais contrafactuais*. O seguinte exemplo é canônico na literatura sobre o tema. Considere as seguintes sentenças: "Se Shakespeare não escreveu *Hamlet*, outra pessoa escreveu", e "Se Shakespeare não tivesse escrito *Hamlet*, alguém teria escrito". Embora parecidas, tais sentenças condicionais possuem valores de verdade bastante distintos: enquanto a primeira é facilmente aceita, é improvável que o leitor encontre alguma pessoa (até mesmo um filósofo) disposta a aceitar a segunda.

A grande revolução no estudo dos condicionais ocorreu no final do século passado, principalmente impulsionado pelos trabalhos de Stalnaker (1984) e Lewis (1973). O desafio era lidar precisamente com casos semelhantes ao do parágrafo anterior. Resumidamente, a estratégia foi criar uma classe para condicionais contrafactuais onde o antecedente é tomado como um caso hipotético, como em “Se Shakespeare não tivesse escrito *Hamlet*, alguém teria escrito”. A principal motivação dos autores era definir condições de verdade para sentenças condicionais gerais. E isso foi feito inicialmente por Lewis (1973), e posteriormente ampliado por Stalnaker (1984), a partir de uma semântica de mundos possíveis (muito semelhante à semântica de esferas que vimos na Seção 4.2.1).

Um outro paradigma no estudo de contrafactuais é o que ficou conhecido como *Teste de Ramsey*. Frank Ramsey (1931) tinha a motivação epistêmica de entender o que seria possuir uma crença condicional. Em uma famosa nota de rodapé o autor escreve:

> Se duas pessoas estão discutindo ‘Se p ocorrerá q?’ e ambas estão em dúvida sobre p, elas estão adicionando p hipoteticamente em sua caixa de conhecimento e discutindo sobre q a partir de tal base [...] Podemos dizer que elas estão fixando seus graus de crença em q dado p. (RAMSEY, F., 1931, p. 249)

Não é difícil notar uma correlação entre a passagem acima e o operador de revisão de crença $*$ que vimos anteriormente. Informalmente, em termos de revisão de crença, o Teste de Ramsey sustenta que dizer que “Se E, então A” faz parte do conjunto de crenças de alguém é o mesmo que dizer que tal pessoa está disposta a aceitar A quando receber E como uma nova informação.

Para ser mais preciso, vamos denotar aqui o condicional “geral”[21] expresso na sentença “Se E, então A” como “$E \Rightarrow A$”. Assim, podemos formular o teste de Ramsey como a seguinte condição:

$$E \Rightarrow A \in \mathbf{B} \textit{ sse } A \in \mathbf{B}^*_E.$$

Ou utilizando nossa notação usual:

$$Bel(E \Rightarrow A) \textit{ sse } Bel(A|E).$$

onde $Bel(\cdot|\cdot)$ representa a operação de revisão qualitativa, seja ela uma operação na teoria AGM ou não-monotônica, como visto na Seção 4.2.3.

[21] Também conhecido como “condicional plano”, do inglês *flat conditional*.

Finalmente, podemos dizer que tal condicional "geral" é simplesmente o representante formal da supracitada operação de mudança de visão qualitativa. Isso significa que, se no momento inicial o agente possui a crença condicional $Bel(A|E)$, e no momento seguinte ele recebe a informação de que a proposição E é o caso, então é diacronicamente racional acreditar em A no momento seguinte. Mais precisamente, sendo t um instante anterior a t', se entre t e t' o agente recebe a evidência E, então

$$Bel_{t'}(A) \textit{ sse } Bel_t(A|E) \textit{ sse } Bel_t(E \Rightarrow A).$$

Para finalizar, é importante reforçar que não é nosso objetivo aqui se aprofundar na rica literatura sobre condicionais. Meu objetivo é apenas utilizar da noção de "crença condicional" para caracterizar melhor o que venho chamando até aqui de "mudança de visão". E isso implica na possibilidade de traçar paralelos entre os modelos formais de revisão qualitativa e quantitativa para racionalidade diacrônica. Em outras palavras, o princípio bayesiano de condicionalização e os modelos de revisão, AGM ou não-monotônico, compartilham de uma literatura em comum que motivou boa parte das referências utilizadas no presente capítulo. Visando a noção de um PP diacrônico, meu objetivo daqui em diante será investigar como crenças revisadas qualitativamente por uma evidência E se relacionam com uma distribuição de probabilidade atualizada por E, a partir do princípio de condicionalização bayesiana. Para isso, vamos retomar o supracitado Paradoxo da Loteria, mas agora por outra perspectiva.

4.3.1 O Paradoxo da Loteria: O Desafio Agora é Outro

No primeiro capítulo vimos como o Paradoxo da Loteria representa um obstáculo para a construção de um PP consistente e não-cético. No terceiro capítulo, apresentamos duas alternativas a um PP que pretende ser consistente e não-cético a partir da Teoria Probalógica (*PL*) e da Tese Humeana (*TH*). Vimos que as duas teorias solucionam o Paradoxo da Loteria tal como o conhecemos. Todavia, até então vimos apenas os paradoxos para os casos sincrônicos. Agora veremos como podemos expandir o desafio lotérico para casos diacrônicos, ou seja, quando há mudança de visão.

O exemplo a seguir foi originalmente oferecido pela dupla Lin e Kelly (2012b, p. 958). Suponha uma loteria com 3 bilhetes, A, B, C, e um limiar lockeano $r \geq 3/4$. Nesse caso não há paradoxo no caso sincrônico, pois $P(\neg A) = 2/3 < 3/4$. Mas suponha agora que a loteria não seja justa, e o bilhete A ganha com probabilidade $1/2$, e os bilhetes B e C com probabilidade $1/4$. Como a probabilidade de derrota dos bilhetes B e C está dentro do limiar de aceitação, podemos acreditar que A vence. Agora considere o caso onde uma nova informação é adquirida pelo apostador no momento seguinte, afirmando que o bilhete C foi removido da loteria, o que implica que sua probabilidade passou a ser 0. Essa informação deveria reforçar a vitória do bilhete A. Mas após a remoção do bilhete C a probabilidade de A passou a ser $2/3$ e $P(B) = 1/3$, de modo que a razão entre eles permanece 2:1. Mas agora o limiar lockeano não permite mais a crença na vitória de A, visto que sua probabilidade de vitória, ou a probabilidade de derrota do bilhete B, não é suficiente para aceitação.

O propósito de tal exemplo é mostrar que um PP gerado por meio da Tese Lockeana falha em satisfazer o seguinte princípio diacrônico: crenças previamente aceitas não devem ser retraídas quando uma de suas consequências lógicas é aprendida. Sendo mais específico, esse princípio é conhecido na literatura como *Princípio do Dedutivismo Hipotético*, e pode ser formulado nos seguintes termos:

DEFINIÇÃO 29 (*Dedutivismo Hipotético*)**.** Se $Bel(A)$, e $A \models E$, então $Bel(A|E)$.

Este é mais fraco que o princípio (*KLM*5) de Monotonicidade Cautelosa, visto na seção 4.2.2, ou seja, segue-se logicamente dele. Isso quer dizer que um modelo de mudança de visão que safisfaz ao menos Monotonicidade Cautelosa, satisfaz o Dedutivismo Hipotético. E isso é o caso tanto para a teoria AGM quanto para o Sistema Preferencial não-monotônico.

Pensemos agora em como a regra lockeana viola tal princípio, como vimos no exemplo lotérico acima. Seja $[\mathbf{B}]_P$ o conjunto de crenças gerado, a partir de *TL*, por uma distribuição de probabilidade subjetiva P, e $[\mathbf{B}]_{P|E}$ o conjunto de crenças gerado pela mesma distribuição P condicionalizada por uma evidência E. Temos que, se $Bel_P(A)$ e $A \models E$, então $Bel_{P|E}(A)$. Uma vez que o Dedutivismo Hipotético é um postulado para a teoria AGM e para o Sistema Preferencial, temos que uma versão diacrônica de *TL* é incompatível com qualquer um dos nossos modelos formais para revisão de crença qualitativa.

Mas ainda podemos ir um pouco mais longe ao analisar *TL* aplicada diacronicamente ao Paradoxo da Loteria. Considere agora o seguinte exemplo, também retirado de Lin e Kelly (2012b, p. 960). Suponha agora que adotemos um PP que possui um limiar dependente do conteúdo, com valores para disjunções mais alto do que para conjunções, de modo que a aceitação de A seja fechada para condicionalização sobre $\neg C$, ou seja, $P(A|\neg C) \geq r$. Nesse caso, ao remover o bilhete C de nossa loteria enviesada, a crença na vitória do bilhete A permanece, satisfazendo assim o Dedutivismo Hipotético. Agora suponha uma distribuição de probabilidade onde a probabilidade de $A \cup B$ esteja acima do limiar r, e próximo, mas não o suficiente, da aceitação de A (isso pode ser observado na figura 13**(b)** que veremos a seguir). Como $P(A)$ está muito próxima do limiar de aceitação para o bilhete A, se em t_2 condicionalizarmos $P(A)$ sobre a derrota de C, ou seja, $P(A|\neg C)$, isso faz com que A seja aceita na nova distribuição de probabilidade, uma vez que o espaço amostral foi reduzido o suficiente para tornar $P(A) \geq r$. Pelo fechamento lógico, concluímos que $\neg B$. Por outro lado, se em t_2 condicionalizarmos nossa distribuição de probabilidade sobre a vitória de C, é certo que B não é vencedor. Logo, independentemente do que ocorra com C, $\neg B$ será aceita. Todavia, antes da condicionalização, o agente não acreditava na derrota do bilhete B, isto é, $\neg B \notin [\mathbf{B}]_P$.

Mais uma vez, o propósito de tal exemplo é mostrar que um Princípio Ponte gerado por meio de *TL* falha em satisfazer o seguinte princípio diacrônico: se o agente sabe que em t_2 irá aceitar uma proposição independente da evidência adquirida, ele já deve aceitar a proposição em t_1. Mais precisamente, tal princípio é reconhecido na literatura como *Princípio do Raciocínio de Caso*, e pode ser formulado nos seguintes termos:

Definição 30 (*Raciocínio de Caso*)**.** Se $Bel(A|E)$ e $Bel(A|\neg E)$, então $Bel(A)$.

Este, possui forte apelo intuitivo, e é válido tanto no modelo AGM quanto no modelo Preferencial para revisão de crenças. Ele consiste em uma versão análoga do famoso *Princípio de Reflexão* desenvolvido por Van Fraassen (1984) como uma norma diacrônica para probabilidades subjetivas[22]. Formulado em termos de um princípio diacrônico para a Tese Lockeana, temos o seguinte: Se $Bel_{P|E}(A)$ e $Bel_{P|\neg E}(A)$, então $Bel_P(A)$.

Os exemplos acima foram apresentados por Lin e Kelly (2012b) visando uma compreensão mais ampla sobre como um PP apropriado deve ser formulado de modo que satisfaça algumas condições, como as que foram apresentadas. Agora veremos como isso se desenrola a partir da sua visualização geométrica, como fizemos no capítulo anterior, mas agora de um ponto de vista diacrônico.

4.3.2 A Geometria do Paradoxo Diacrônico

Como vimos no capítulo anterior, seção 3.3, podemos ampliar nossa visão sobre a natureza de um PP a partir de sua representação geométrica. Esse tipo de modelagem nos permite visualizar rapidamente alguns elementos preciosos para a construção de um PP adequado, como sua consistência e não-ceticismo. Agora devemos nos perguntar: será que é possível expandir tal recurso para o caso de um PP diacrônico? Nesse tópico veremos como isso é possível, baseado fundamentalmente nos trabalhos de Lin e Kelly (2012b, 2021).

Como anteriormente, representaremos nosso (bastante limitado) universo de possibilidades a partir de um $2-$*simplex* para 3 proposições mutuamente excludente e coletivamente exaustivas. Primeiramente, dado que estamos trabalhando dentro de uma partição, é fácil ver que submeter uma distribuição de probabilidade a uma nova evidência conclusiva impactará decisivamente nosso resultado. Para ser mais objetivo, o caso mais óbvio é quando qualquer elemento da partição é condicionalizado por um elemento diferente dele. Isso implica que a probabilidade do elemento em questão irá sempre à zero após a condicionalização. No caso lotérico, por exemplo, dado que

[22] Tal princípio é, juntamente com o princípio de condicionalização bayesiana, alvo das críticas a uma noção de racionalidade diacrônica, como foi apresentado no início do presente capítulo.

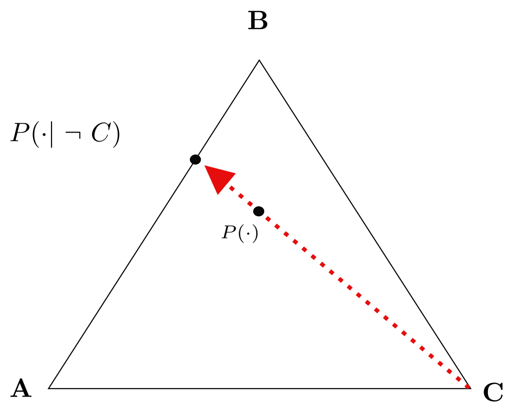

Figura 11 – Uma representação geométrica da condicionalização quando a informação $\neg C$ é adquirida.

o bilhete *B* é o vencedor, a probabilidade de qualquer um dos bilhetes restantes é automaticamente atualizada para zero, quando condicionalizado sobre a vitória de *B*. Um caso menos óbvio, mas ainda bastante simples, é quando condicionalizamos um elemento da partição pela rejeição de algum outro elemento diferente dele. Nesse caso, nossa partição perderá um elemento e, uma vez que as probabilidades são todas independentes entre os elementos da partição, a nova distribuição de probabilidade deverá "distribuir" proporcionalmente a probabilidade do elemento rejeitado para os restantes (pense novamente no caso lotérico onde recebemos a informação de que um bilhete é o perdedor, isso aumenta proporcionalmente a probabilidade dos bilhetes restantes serem vencedores).

Visualmente, em nossa representação com um 2 – *simplex*, temos que uma condicionalização[23] reduz o 2 – *simplex* a um 1 – *simplex*, ou seja, nosso triângulo equilátero passa a ser uma linha unitária, tal como vimos no início da Seção 3.3, no capítulo anterior, na Figura 1. Agora no caso diacrônico, a Figura 11 representa uma condicionalização aplicada à partição quando um de seus elementos é rejeitado. Nesse caso, o ponto representando a distribuição de probabilidade inicial $P(\cdot)$ é "empurrado" para o vértice oposto à proposição excluída, reduzindo o 2–*simplex* para um 1–*simplex*, isto é, o segmento $\overline{AB}$.

Vamos agora pensar em alguns requisitos que um PP diacrônico deve satisfazer olhando para sua representação geométrica. O primeiro deles consiste no fato de que a região de aceitação de cada uma das proposições está relacionada com uma proximidade com os vértices do triângulo. Visualmente, isso significa que a crença em uma resposta *A* em uma distribuição de probabilidade *P* implica que *A* deve ser aceita em cada um dos pontos sobre a reta que conecta *P* ao vértice do triângulo onde *A* possui probabilidade máxima, como podemos observar na Figura 12. O que a figura diz é que,

[23] Vamos trabalhar aqui apenas com condicionalização padrão, isto é, quando a nova informação adquirida possui probabilidade 1. Sobre casos com condicionalização de Jeffrey, ver Lin e Kelly (2021).

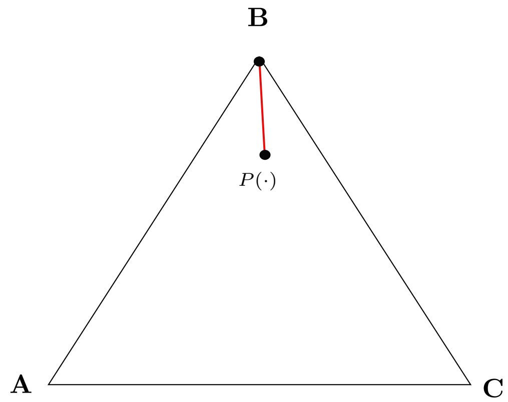

Figura 12 – Monotonicidade de canto.

se B é aceito em $P(\cdot)$, então B também deve ser aceito em qualquer outra distribuição P' que se encontre sobre a linha vermelha conectada ao vértice de B. Tal exigência geométrica representa o seguinte requisito para um PP diacrônico: um agente racional não deve abandonar sua crença em A quando a probabilidade subjetiva de A aumentar após condicionalização. Lin e Kelly (2012b, p. 965) chamam essa exigência de *Monotonicidade de Canto*, ou também *Monotonicidade de Correspondência* em Lin e Kelly (2021, p. 198). Tal requisito é relativamente simples de ser satisfeito e, como veremos a seguir, mesmo *TL* o satisfaz.

O segundo requisito que um PP diacrônico deve satisfazer foi apresentado na seção anterior, o Dedutivismo Hipotético. Agora olhemos para sua interpretação geométrica. Na Figura 13 (*a*) vemos que o problema diacrônico para *TL*, conforme apresentado no tópico anterior, possui uma razão geométrica: após condicionalização, a distribuição de probabilidade inicial $P(\cdot)$ que aceitava a proposição B por meio de *TL* foi conduzida a uma nova distribuição $P(\cdot \mid \neg C)$ onde B não é mais aceita em *TL*, mesmo $\neg C$ sendo uma consequência lógica de B. Isso significa que a forma geométrica de *TL* possui regiões de aceitação demasiadamente "pontiagudas" nas proximidades das arestas, e isso implica que o Dedutivismo Hipotético não é satisfeito, como visto anteriormente.

O terceiro requisito, também apresentado na seção anterior, é o princípio de Raciocínio de Caso. Olhando para a motivação geométrica, na Figura 13 (*b*) vemos um caso onde *TL* possui limiares distintos dependendo da proposição em questão. Lin e Kelly (2021) chamam tal versão lockeana de "Lockeanismo dependente do conteúdo". No exemplo da Figura 13 (*b*), o limiar de aceitação para disjunções é mais alto do que o limiar para conjunções (recordemos que a zona de aceitação de B consiste na intersecção entre a zona de aceitação de $\neg C$ e $\neg$ A). Nesse caso, o Dedutivismo Hipotético é satisfeito, uma vez que a crença em B está fechada para condicionalização sobre $\neg C$. Todavia, temos que a distribuição de probabilidade gerada por *TL* em $P(\cdot \mid \neg C)$ acarreta $\neg A$ e a distribuição gerada por $P(\cdot \mid C)$ também acarreta, automaticamente, $\neg A$, mas a distribuição inicial $P(\cdot)$ não acarreta $\neg A$. Consequentemente, o requisito de

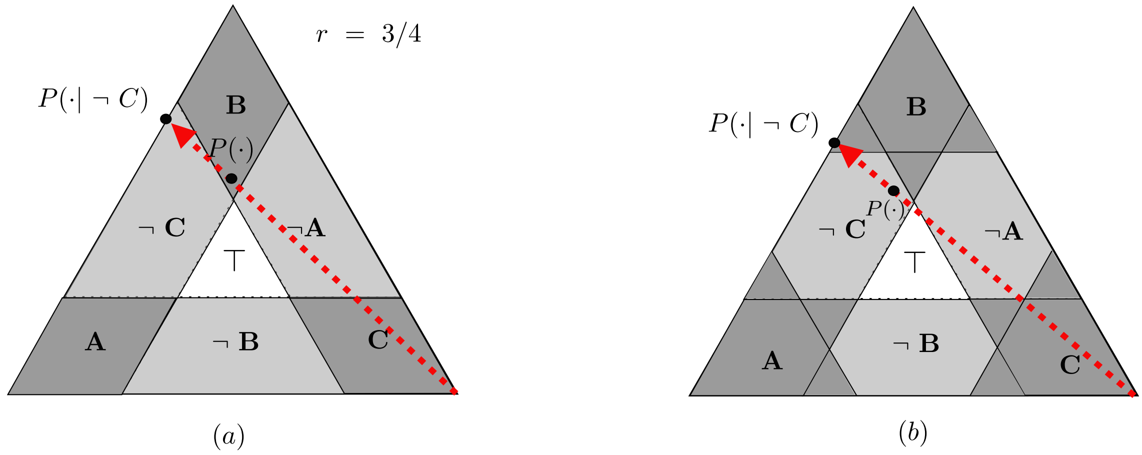

Figura 13 – Os dois novos desafios à Tese Lockeana.

Raciocínio de Caso não é violado.

Com isso, visualizamos novamente a geometria de *TL*, mas agora de um ponto de vista diacrônico. Enquanto a teoria satisfaz facilmente Monotonicidade de Canto, *TL* não satisfaz o Dedutivismo Hipotético, e mesmo uma versão lockeana com limiar dependente do conteúdo falhará em satisfazer o requisito de Raciocínio de Caso. Isso significa que a Tese Lockeana não é limitada apenas pelo desafios de consistência amplamente mencionados aqui desde o início, mas também por exigências diacrônicas. Consequentemente, se desejamos encontrar um PP capaz de atuar como um requisito racional durante uma mudança de visão, precisamos buscar por alternativas.

4.3.2.1 Em busca de um PP diacrônico para além de *TL*

Vamos agora dar um passo além e expandir a investigação para as duas alternativas teóricas de PP que apresentamos no capítulo anterior, a Tese Humeana de Leitgeb e a Teoria Probalógica de Lin e Kelly. Vamos inicialmente comparar a interpretação geométrica de ambas (já vistas no capítulo anterior, Seção 3.3) no que diz respeito a condicionalização bayesiana e revisão de crença. Na Figura 14, interpomos as duas teorias com limiares equivalentes para compará-las geometricamente. É fácil vizualizar como ambas satisfazem o requisito de Monotonicidade de Canto. Ainda, podemos notar que tanto $TH_{\mathcal{Y}}$ quanto *PL* satisfazem ambos os requisitos de Dedutivismo Hipotético e Raciocínio de Caso.

Ainda olhando para a Figura 14, como podemos observar, a área de cor amarela consiste precisamente na região do triângulo onde humeanos são céticos e probalógicos são opinados. Como vimos, o fato de a zona cética de $TH_{\mathcal{Y}}$ ser bem maior que de *PL*, torna $TH_{\mathcal{Y}}$ uma opção de PP mais cética, dado limiares equivalentes. Mas isso já foi abordado anteriormente. Estamos interessados agora em comparar diacronicamente tais opções de PP.

A Figura 15 mostra um caso de condicionalização aplicado em ambas teorias.

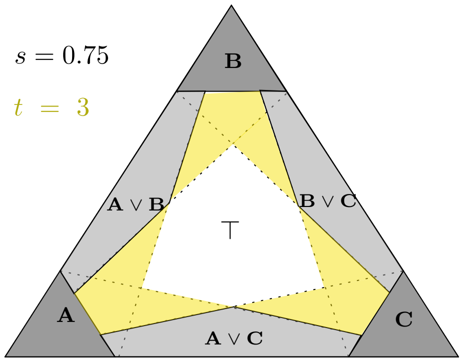

Figura 14 – Interpondo a geometria de $TH_{\mathcal{Y}}$ com PL.

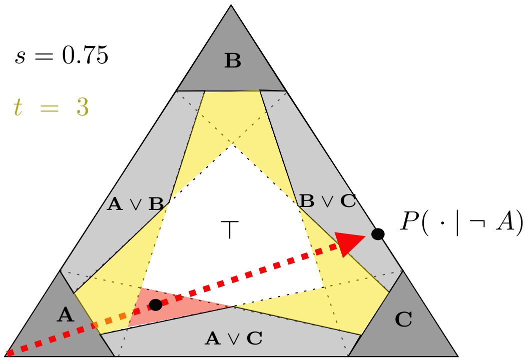

Figura 15 – Condicionalização sobre uma mesma distribuição de probabilidade, dada uma nova evidência $\neg A$.

Estou aqui interessado especificamente nas pequenas áreas onde PL e $TH_{\mathcal{Y}}$ divergem, como a região destacada em vermelho na figura. Tal área é relevante principalmente na medida em que queremos entender o que ocorre nessa região durante uma condicionalização. Precisamente sobre a região em vermelho, uma distribuição de probabilidade, em qualquer ponto dessa região, implica que PL acredita na proposição $A \vee C$, enquanto $TH_{\mathcal{Y}}$ possui em seu conjunto de crenças somente a proposição tautológica, isto é, $A \vee B \vee C$. Mas o que ocorre com os respectivos conjuntos de crença quando o recebemos a informação de que $\neg A$ é o caso?

Como podemos ver, durante a condicionalização por $\neg A$, uma consequência de $TH_{\mathcal{Y}}$ é que o agente terminará aceitando a proposição $B \vee C$. E isso soa bastante lógico: quando $\neg A$ é aprendido, apenas removemos A do conjunto inicial $A \vee B \vee C$, e passamos a acreditar apenas em $B \vee C$. Todavia, PL não funciona da mesma forma. O conjunto inicial de crenças gerado pela distribuição $P(\cdot)$, localizada no interior da região destacada em vermelho, é $A \vee C$. Uma vez que estamos trabalhando com uma partição de elementos disjuntos, como previsto, $A \vee C$ é logicamente equivalente a $\neg B$. Como nosso conjunto de crenças é fechado logicamente, temos que $Bel(\neg B)$. Todavia, quando $\neg A$ é adquirido como nova informação, $\neg B$ é removido de nosso conjunto de crenças, uma vez que o agente termina acreditando em $B \vee C$, simplesmente. Esse simples fato geométrico distinguindo a dinâmica de revisão entre $TH_{\mathcal{Y}}$ e PL decorre de uma divergência profunda entre tais teorias no que diz respeito ao método de revisão de crenças.

Resumidamente, o que podemos extrair do exemplo acima é que *PL* não satisfaz a seguinte norma de revisão: Se A é compatível com o conjunto inicial de crenças, ou seja, não é o caso que $Bel(\neg A)$, então $Bel(B \mid A)$ se, e somente se, $Cn([\mathbf{B}] \cup \{A\}) \models B$. Resumidamente, tal norma diz que não há contração sem refutação. O leitor mais familiarizado deve ter notado que tal condição é basicamente a união dos postulados ($B^{*}3$) e ($B^{*}4$) da teoria AGM, e uma consequência do princípio (*KLM*8) de Monotonicidade Racional em um modelo Preferencial não-monotônico de revisão de crença. Todavia, para Lin e Kelly (2012b), o fato de *PL* não satisfazer tal norma é justamente uma de suas virtudes. Para os autores, tal condição é impossível de ser satisfeita para qualquer modelo formal de revisão de crença que vise *rastrear* o princípio de condicionalização bayesiana. Isso significa que, no que diz respeito ao método dinâmico de revisão de crença, $TH_{\mathcal{Y}}$ e *PL* possuem divergências profundas. O restante da presente Tese será sobre tal divergência, como veremos a seguir.

4.3.3 Estabilidade *Vs.* Rastreio

Mais do que uma diferença superficial, a diferença entre $TH_{\mathcal{Y}}$ e *PL* representada pela zona de aceitação destacada cima, conforme vimos na Figura 15, demarca o modo como devemos compreender o papel de um Princípio Ponte na discussão sobre a racionalidade da crença e probabilidade subjetiva. Como veremos, por um lado $TH_{\mathcal{Y}}$ sustenta que a estabilidade de nossas crenças é um atributo fundamental de nossa racionalidade, mesmo que isso nos custe a impossibilidade de uma correspondência perfeita (biunívoca) entre a dinâmica da crença e da probabilidade subjetiva. Por outro lado, *PL* sustenta que é somente abrindo mão de conjuntos de crenças estáveis que seremos capazes de determinar um conjunto de crenças exclusivamente em função do nosso conjunto de probabilidades subjetivas, em outras palavras, reduzir uma racionalidade à outra. Como veremos, o segundo caso nos oferece um modelo perfeito para a redução da racionalidade de crenças em termos de probabilidades subjetivas, e o primeiro nos oferece uma visão não-reducionista onde não existe uma hierarquia entre a racionalidade das duas noções relevantes, mas apenas uma norma de coerência.

4.3.3.1 Rastreio

O conceito de "rastreamento" será central no presente tópico. Para definir melhor essa noção precisamos retomar alguns conceitos já previamente vistos. Primeiramente, para cada distribuição de probabilidade P sobre W, dizemos que $[\mathbf{B}]_P$ é o conjunto de crenças gerado por meio de um determinado PP, dada uma distribuição P. Cada PP abordado até aqui pode gerar conjuntos de crenças $[\mathbf{B}]_P$ distintos, dada a mesma distribuição P. E podemos notar facilmente isso olhando para nosso 2 – *simplex* e as regiões de aceitação para cada PP relevante. Agora, avançando um pouco mais, dizemos que uma teoria formal de revisão de crença *rastreia* a condicionalização bayesiana se, e somente se, quando revisado por uma nova evidência E, o conjunto de crença $[\mathbf{B}]_{P|E}$ é igual ao conjunto $[\mathbf{B}]_P$ revisador por E em uma operação de revisão qualitativa, onde $P|E$ é a distribuição de probabilidade condicionalizada sobre E. Mais precisamente:

Definição 31 (Rastreamento)**.** Para todo $A_i \in \mathcal{A}$, $P(E) > 0$, e toda distribuição $P \in \mathcal{P}$, temos que:

$$Bel_P(A_i \mid E) \text{ sse } Bel_{P|E}(A_i)$$

Ou de modo equivalente:

$$[\mathbf{B}]_P * E = [\mathbf{B}]_{P|E}$$

Em outras palavras, a condição de rastreamento afirma que a rota para uma atualização de probabilidades subjetivas por uma evidência E, que dá origem a um conjunto de crenças [**B**] por meio de um PP, deve levar ao mesmo lugar que a revisão de crença qualitativa em [**B**] quando E surge como uma nova informação. Isso significa que existe uma correspondência entre a condicionalização bayesiana e o modelos de revisão de crença qualitativa. Quando um PP específico possui como modelo uma teoria formal de revisão que satisfaz a condição acima, dizemos que tal princípio ponte *rastreia a condicionalização bayesiana*.

Uma das condições para qualquer PP que deseje rastrear condicionalização é que P e [**B**] possuam uma correspondência bi-unívoca, ou seja, que exista uma bijeção entre o conjunto $\mathcal{P}$ de todas distribuições P possíveis, e a coleção dos possíveis conjuntos de crença [**B**]. Para Lin e Kelly (2021, p. 186), *PL* é capaz de fornecer o que eles chamam de *correspondência doxástica*: uma mapeamento de probabilidades subjetivas para crenças. Isso significa que, nesse caso, [**B**] é determinado integralmente em função de P. Ou seja, um PP é basicamente uma função f de aceitação cujo argumento é uma distribuição de probabilidade P, e a saída um conjunto de crenças $[\mathbf{B}]_P$, mais precisamente,

$$f(P) = [\mathbf{B}]_P.$$

Como veremos abaixo, tal *correspondência doxástica* em um PP só será possível na medida em que rejeitamos qualquer modelo de revisão qualitativa de crença que satisfaz o requisito de Monotonicidade Racional, ou sua dupla equivalente (B^*3) e (B^*4) na teoria AGM.

Vamos então nos voltar agora para as condições que um modelo de revisão qualitativa de crença deve satisfazer e as implicações disso para a construção de um PP adequado. Primeiramente, para Lin e Kelly (2012b, p.964) um PP compatível com a Monotonicidade Racional, ou também (B^*3) e (B^*4), é dito *cumulativo*, ou *acretivo*: proposições consistentes com o que o agente acredita podem ser acumuladas e servirem de premissas para inferências futuras. Não surpreendentemente, a teoria AGM é uma teoria cumulativa. Isso significa que qualquer PP diacrônico que possua como modelo de revisão a teoria AGM será, por consequência, um PP cumulativo.

Em segundo lugar, para os autores (2012, p. 965), um PP deve satisfazer as seguintes condições:

1. ser não-cético (ou seja, permite crença com probabilidade menor que 1);
2. não-opinado (ou seja, permite crenças incompletas ou disjuntivas);
3. consistente (isto é, não existe uma distribuição de probabilidade P tal que $Bel_P(\bot)$); e
4. monótono de canto (isto é, se $Bel_P(A)$ e temos que $P(A \mid E) > P(A)$, então $Bel_{P|E}(A)$).

Um PP é dito *sensível* se, e somente se, satisfaz todas as condições acima. Um dos principais resultados do trabalho de Lin e Kelly foi mostrar que um PP *sensível* que *rastreia condicionalização* não pode ser *cumulativo*, isto é, ser compatível com a Monotonicidade Racional. Esse resultado é consagrado a partir do seguinte teorema:

TEOREMA 4.3.1. *Se $\mathcal{Q}$ é uma questão com pelo menos 3 respostas, nenhum PP sensível que rastreia condicionalização bayesiana é cumulativo.*

Os detalhes desse resultado podem ser conferidos no Apêndice B. Como corolário, qualquer PP sensível que possua como modelo de revisão a teoria AGM, não pode rastrear a condicionalização bayesiana, e consequentemente, não pode estabelecer uma correspondência doxástica entre o conjunto de crenças e o conjunto de probabilidades subjetivas.

Uma das motivações dos autores para rejeitar cumulatividade é baseada em um famoso exemplo na história da filosofia: uma variação dos contraexemplos de Gettier, mais especificamente, o caso "sem lemas falsos". O exemplo é descrito da seguinte maneira (LIN; KELLY, 2012b, p. 964):

> Você está preocupado se dois de seus conhecidos, Nogot e Havit, possuem um carro Ford específico. Nossa partição consiste em três possibilidades: (i) "Nogot tem um Ford"; (ii) "Havit tem um Ford"; e (iii) "nenhum deles tem um Ford". Você possui certa evidência de que Nogot possui um Ford - viu ele certo dia dirigindo um, e apenas um indício de que Havit possui um Ford. Embora você não consiga dizer qual deles possui um Ford, suas evidências são suficientes para acreditar que algum deles possui um Ford. Ao questionar Nogot, ele lhe conta que o carro que você viu ele dirigindo era alugado, o que derrota a parte mais forte de sustentação para sua

> crença de que algum deles possui um Ford. Portanto, você abandona sua crença inicial de que algum deles possui um Ford, mesmo que a evidência adquirida, “Nogot não tem um Ford”, era logicamente consistente com a crença inicial de que algum deles possuia um Ford.

O exemplo acima é tratado pelos autores como um caso onde Monotonicidade Racional é violada, e isso seria uma boa razão para rejeitarmos uma teoria de revisão cumulativa.

Vamos explorar um pouco mais detalhadamente o caso acima de modo que sua motivação e conclusão tornem-se suficientemente claras. Tome a proposição A como denotando a afirmação “Nogot tem um Ford”, B como “Havit tem um Ford” e C afirmando “nenhuma deles tem um Ford”. Em termos de um ranqueamento de plausibilidade, podemos afirmar que A é mais plausível do que C, mas B não é mais plausível do que C. Dado a crença inicial em $A \cup B$ e uma ordem de plausibilidade, a nova informação $\neg A$ torna C mais plausível, de tal modo que o novo conjunto de crenças será $B \cup C$. Para os autores, tal raciocínio parece correto, tanto para eles, quanto para boa parte da comunidade epistemológica que discute os casos Gettier (LIN; KELLY, 2021, p. 193). Por outro lado, uma revisão AGM terminaria de outro modo: uma vez que o agente acredita em $A \cup B$, C não pode ser mais plausível do que A nem do que B. Uma vez que o modelo cumulativo AGM implica uma ordem total (como vimos na Seção 4.2.1), C só pode estar ranqueada *abaixo* tanto de A quanto de B. Isso faz com que C continue sendo rejeitada, mesmo quando A é removida, o que faz com o que o agente termine acreditando em B[24].

Para Lin e Kelly, a única estratégia eficiente para um PP dar conta de casos, semelhantes aos acima apresentados, é levando em conta a proporcionalidade, ou cotações, em sua construção (tal como vimos no capítulo anterior, Seção 3.2). A estratégia adotada por eles é precisamente uma teoria da revisão de crença não-monotônica a partir de um modelo Preferencial ao estilo de Shoham (1987), já exploradno anteriormente aqui na Seção 4.2.2. Relembrando, começamos por uma ordem bem fundada (ou seja, não possui uma cadeia infinita descendente), parcial estrita (irreflexiva e transitiva), sobre o conjunto de respostas completas para $\mathcal{Q}$. Ou seja, uma ordem de plausibilidade onde $w_i \prec w_j$ significa que w_i é estritamente mais plausível do que w_j, onde $w_i, w_j \in W$ formam uma partição de respostas completas para uma questão $\mathcal{Q}$. Uma ordem de plausibilidade induz um método de revisão de crença de um Sistema Preferencial: Dada uma informação E, o novo conjunto de crenças [**B**] será a disjunção das respostas mais plausíveis que são compatíveis com E, gerando assim uma nova ordem de plausibilidade $\prec_{|E}$, ou também $\prec^*_E$, e por fim basta tomar as repostas mais plausíveis gerada pela nova ordem de plausibilidade $\prec^*_E$ para obter o novo conjunto [**B**].

[24] Geometricamente, esse é precisamente o caso da Figura 15.

Para Lin e Kelly (2012b, p. 968), um novo PP gerado a partir de uma ordem de plausibilidade $\prec$ e uma distribuição de probabilidade P pode ser definido em termos de cotações (*odds*) como:

Definição 32. Sejam w_i e w_j respostas completas para $\mathcal{Q}$, com $P(w_j) > 0$, e t um limiar com valor estritamente maior que 1:

$$w_i \prec_P w_j \; sse \; \frac{P(w_i)}{P(w_j)} > t$$

Aqui, $\prec_P$ é uma ordem de plausibilidade determinada por uma distribuição particular $P \in \mathcal{P}$. E como vimos no capítulo anterior, na definição 23, o lado direito da relação acima determinará por meio de PL se uma proposição será ou não parte do conjunto de crenças do agente. Conforme Lin e Kelly (2012b, p. 967), comparando cada par w_i, w_j de respostas completas para $\mathcal{Q}$, dada uma distribuição de probabilidade P, chegamos ao nosso conjunto relevante de crenças $[\mathbf{B}]_P$, que é exatamente o mesmo gerado por PL, mas agora partindo de uma ordem de plausibilidade $\prec_P$.

Para facilitar nossa compreensão, vamos ao seguinte exemplo. Tome $t = 3$ na definição acima. Esse valor de limiar dará origem precisamente ao 2 – *simplex* das Figuras 14 e 15. Relembrando, o lugar geométrico onde $\frac{P(A)}{P(B)} = 3$ é o segmento inferior com origem no vértice C que intersecciona a aresta $\overline{AB}$, mais especificamente, a *linha de cotações constantes*, conforme vimos no Capítulo 3, Seção 3.3.1. Para determinar se $A \prec_P B$, simplesmente observe se a distribuição de probabilidade P está abaixo do segmento de reta em questão. Faça isso com cada par de resposta e teremos precisamente a região de aceitação relevante dentro de nosso 2 – *simplex* a partir de PL com $t = 3$.

A partir da ordem de plausibilidade da Definição 32, agora podemos determinar, dada uma distribuição de probabilidade P, o conjunto relevante de crenças, do mesmo modo que vimos no capítulo 3:

Definição 33 (Conjunto de crenças em PL)**.** Seja $t > 1$, e $w_i, w_j \in W$ respostas para $\mathcal{Q}$:

$$Bel(A) \; sse \; A \supseteq \left\{ W \setminus w_i : \frac{P(w_i)}{\max_j P(w_j)} \leq \frac{1}{t} \right\}$$

O diferencial agora é que a definição acima irá dar conta não apenas da racionalidade sincrônica do conjunto de crenças, mas também do processo diacrônico, por meio de uma ordem de plausibilidade em um Sistema Preferencial não-monotônico.

Uma das grandes vantagens em se utilizar uma ordem de plausibilidade para revisão de crença é que ela nos oferece o que Lin e Kelly (2012, p. 961) chamam de "Espaço Proposicional das Razões". Isso significa que, dado um modelo de revisão qualitativa de crença, podemos raciocinar diretamente com as proposições e sua ordem de plausibilidade, sem ter que consultar constantemente as probabilidade subjecentes.

Mas para que isso possa ocorrer, é necessário satisfazer a correspondência doxástica entre crença e probabilidade subjetiva, garantindo que, tanto uma revisão qualitativa por um Sistema Preferencial, quanto uma revisão quantivativa pela condicionalização bayesiana, conduzam sempre ao mesmo lugar. Em outras palavras, que nosso PP em questão rastreie condicionalização bayesiana.

Para finalizar, apresento um compilado dos resultados mais importantes para a teoria Probalógica diacrônica. O teorema a seguir explica grande parte das condições afirmadas acima:

TEOREMA 4.3.2. *A teoria Probalógica com um método de revisão de crença baseado em uma ordem de plausibilidade em um Sistema Preferencial, conforme Definição 32, cumpre os seguintes requisitos:*

i é sensível (se $\mathcal{Q}$ contém ao menos três respostas completas);

ii rastreia condicionalização bayesiana; e

iii satisfaz Dedutivismo Hipotético e Raciocínio de Caso.

Uma versão simplificada da demonstração desse teorema aparece no Apêndice B, onde apresento alguns dos resultados demonstrados em Lin e Kelly (2012b, p. 978).

De um modo mais simples, é possível visualizar geometricamente parte do resultado acima. Para ver como PL é sensível, basta observar a representação geométrica de PL em nosso $2-$ *simplex* e notar que: (i) crenças são aceitas em regiões em que $P(A_i) < 1$; (ii) existem regiões onde crenças disjuntivas (incompletas) são aceitas; (iii) não existe uma região tal que $Bel_P(\perp)$; e, (iv) notoriamente, a monotonicidade de canto é respeitada.

Do mesmo modo, é possível ver que um PP que possui como teoria de revisão de crença o Modelo Preferencial com ordem de plausibilidade, conforme a Definição 32, rastreia condicionalização bayesiana. Para isso, considere uma distribuição P tal que PL implique $Bel(A \vee C)$, tal como vimos na Figura 15, quando $t = 3$. Agora, seguindo o exemplo, considere a nova evidência $\neg A$. Para mostrar que $[\mathbf{B}]_P * \neg A = [\mathbf{B}]_{P|\neg A}$ basta observar a ordem de plausibilidade $\prec_P$. Primeiramente com a revisão qualitativa, no exemplo temos que: $A \prec_P B$, mas $C \not\prec_P B$, e também $C \not\prec_P A$. Agora, vamos restringir a ordem de plausibilidade para o caso onde recebemos a nova informação $\neg A$. A ordem resultante será tal que B e C possuem o mesmo ranqueamento, e não será o caso que $Bel(C)$[25]. Não por acaso, do lado quantitativo, ao condicionalizar $P(B)$ e $P(C)$ sobre a informação $\neg A$, teremos uma ordem de plausibilidade gerada por $\prec_{P|\neg A}$, e podemos

[25] Para ver como isso é o caso, precisamos voltar na Seção 4.2.2, e retomar a definição 27, que diz que $\neg A \mid\!\sim_{[\mathbf{B}]_P} C$ *sse* C é verdadeira em todo mundo $\neg A-$ *minimal*. Mas ao retomar a noção de $\alpha-$*minimal*, percebemos que tanto B quanto C representam, em nossa partição, mundos mutuamente excludentes onde ambos são $\neg A-$ *minimais*. Logo, como C não é verdadeira em mundo exclusivo de B, e tal mundo é $\neg A-$ *minimal*, concluímos que não é o caso que $\neg A \mid\!\sim_{[\mathbf{B}]_P} C$.

ver que não é o caso que $C \prec_{P|\neg A} B$, pois, visivelmente, $\frac{P(C)}{P(B)} \leq 3$, e o agente termina apenas com $Bel(B \vee C)$. Ou seja, tanto $Bel_P(A \vee C \mid \neg A)$ quanto $[\mathbf{B}]_{P|\neg A}$ concluem na mesma crença: $Bel(B \vee C)$.

Finalmente, *PL* satisfaz Dedutivismo Hipotético e Raciocínio de Caso simplesmente pelo fato de que $\prec_P$ é um caso especial do Modelo Preferencial visto anteriormente. Assim, basta notar que Dedutivismo Hipotético e Raciocínio de Caso são consequências do sistema **P**, mais precisamente, dos postulados (*KLM*5) e (*KLM*6).

4.3.3.2 Estabilidade

Para visualizar melhor o que está acontecendo aqui, recomendo que o leitor retome o exemplo da Figura 15. O que está em disputa é exatamente a área destacada em vermelho na imagem. Relembrando, enquanto $TH_{\mathcal{Y}}$ aceita apenas a proposição tautológica (cética) nessa região, *PL* aceita $A \vee C$, que é logicamente equivalente à $\neg B$. No tópico acima vimos que, para a região em destaque na figura ser uma área de aceitação não-cética, e continuar satisfazendo os requisitos sincrônicos para o conjunto de crenças, devemos abrir mão do princípio diacrônico de Monotonicidade Racional, em um modelo não-monotônico de revisão de crença. Veremos agora uma forma de aceitar Monotonicidade Racional como um requisito racional da revisão de crença aceitando apenas a proposição cética na região destacada na figura. Como já mencionamos anteriormente, a vantagem disso é que nosso conjunto de crenças será estável, e possuirá um funcionamento bastante intuitivo: dado a crença em $A \vee B \vee C$, ao receber $\neg A$ como uma nova informação, meu novo conjunto será $B \vee C$. Qual o preço de tal estabilidade? A ausência de um PP capaz de rastrear a condicionalização bayesiana.

Antes de qualquer coisa, vamos relembrar alguns dos elementos centrais de $TH_{\mathcal{Y}}$ que serão imprescindíveis para compreender o que está em questão. No Capítulo 3, ao apresentar a teoria, vimos que a noção relevante de *P*–estabilidade determinava, para qualquer distribuição de probabilidade $P \in \mathcal{P}$, uma menor proposição B_W, cuja responsabilidade era, fundamentalmente, a de dar origem ao conjunto [**B**] de crenças. Já no presente capítulo, Seção 4.2.1, ao apresentar a semântica da teoria AGM a partir do sistema de esferas de Groove, o leitor pode ter notado alguma semelhança entre B_W e um sistema de esferas. E a semelhança não é mero acaso: em um sistema de esferas, [**B**] corresponde pontualmente ao núcleo B_W, e a coleção de esferas, determinada por uma distribuição *P*, denotada como $\mathbb{S}_P$, é exatamente a família de núcleo de crenças $\mathscr{C}_P$ que vimos no início do terceiro capítulo. Relembrando, enquanto B_W é a intersecção de tudo que o agente acredita em W, B_A é a intersecção de tudo que o agente acredita em um conjunto A, sendo $A \subseteq W$. Agora em termos da semântica em AGM, dada uma coleção de esferas $S_1, S_2, \ldots, S_n \in \mathbb{S}$, se A é subconjunto de W, temos que B_A é a intersecção de A com a menor esfera em $\mathbb{S}$ que possui intersecção não-vazia com A, onde “a menor esfera em $\mathbb{S}$” é determinado por meio meio de uma função de seleção. Como veremos aqui, existe uma forte conexão entre $TH_{\mathcal{Y}}$, núcleos de crença, e sistemas de esferas.

Como ponto de partida, visando facilitar a compreensão, vamos retomar um exemplo do Capítulo 3, Seção 3.1.2. Relembrando, as tabelas 1 e 2, reproduzidas novamente aqui, descrevem, respectivamente, um conjunto de possibilidades sobre 3 proposições não mutuamente excludentes, ou seja, um total de 2^3 mundos possíveis, e uma distribuição de probabilidade P sobre os 8 mundos. Para facilitar a leitura, reproduzo novamente as tabelas mencionadas:

$w_1 = A, B, C$	$w_3 = A, B, \neg C$	$w_5 = \neg A, B, \neg C$	$w_7 = \neg A, B, C$
$w_2 = A, \neg B, C$	$w_4 = A, \neg B, \neg C$	$w_6 = \neg A, \neg B, \neg C$	$w_8 = \neg A, \neg B, C$

$P(\{w_1\}) = 0.4$	$P(\{w_3\}) = 0.14$	$P(\{w_5\}) = 0.02$	$P(\{w_7\}) = 0$
$P(\{w_2\}) = 0.32$	$P(\{w_4\}) = 0.11$	$P(\{w_6\}) = 0.01$	$P(\{w_8\}) = 0$

Em nosso exemplo, vimos que o núcleo de crenças B_W era gerado a partir da intersecção de todas proposições P–estáveis, onde uma proposição A é considerada P–estável se, e somente se, $P(A \mid Y) > 1/2$ para todo $Y \in \mathcal{Y}$ tal que $A \cap Y \neq \emptyset$[26]. No caso da distribuição P na tabela inferior, temos os seguintes candidatos a núcleos de crença: $B_W = \{w_1, w_2\}$ ou $B_W = \{w_1, w_2, w_3, w_4\}$, ou $B_W = \{w_1, \ldots, w_6\}$. Um dos grandes resultados teóricos de Leitgeb foi mostrar que esse conjunto de escolhas possíveis para o núcleo de crenças do agente pode desempenhar precisamente o papel de um sistema de esferas $\mathbb{S}$, de tal modo que:

$$\mathbb{S} = \{\{w_1, w_2\}, \{w_1, w_2, w_3, w_4\}, \{w_1, \ldots, w_6\}\}.$$

Na medida em que a distribuição P da Tabela 1 dá origem a um conjunto de proposições P–estáveis ela também gera uma *pré-ordem total entre mundos* de tal modo que $w_1 \preccurlyeq w_2$ significa que w_1 está ranqueado acima de w_2 no que diz respeito a P–estabilidade. Previsivelmente, tal relação de ordem é entendida em termos de um “ranqueamento de plausibilidade”. Como veremos no Teorema 4.3.3 mais abaixo, $\preccurlyeq$ é exatamente a relação de ordem $\leq_{\mathbb{S}}$ entre mundos (conforme vimos na Seção 4.2.1) em que $\leq_{\mathbb{S}}$ é uma relação totalmente ordenada, transitiva, e limitada (isto é, possui um menor elemento).

É fácil ver como o sistema de esferas $\mathbb{S}$ em nosso exemplo é simplesmente um conjunto “aninhado” de mundos não-vazios: para quaisquer esferas $S_i, S_j \in \mathbb{S}$ temos que $S_i \subseteq S_j$ ou $S_j \subseteq S_i$. Uma vez que $\mathbb{S}$ é uma coleção de conjuntos de mundos S_i, podemos entender a relação de plausibilidade entre mundos do seguinte modo: um mundo w_i é ao menos tão plausível quanto w_j se cada esfera S que inclui w_j também

[26] A definição de P–estabilidade aparece no terceiro capítulo, na Definição 22.

inclui w_i. E, como nossa relação de plausibilidade é um pré-ordenamento, isso permite "empates" entre os mundos, ou seja, a relação $\preccurlyeq$ não é estrita.

Ainda, dado $A \in W$, dizemos que B_A é exatamente a intersecção de A com o menor dos membros de $\mathbb{S}$ compatível com A. Em nosso exemplo, $A = \{w_1, w_2, w_3, w_4\}$, então temos que $B_A = A \cap S_1 = \{w_1, w_2, w_3, w_4,\} \cap \{w_1, w_2\} = \{w_1, w_2\}$. Segundo Leitgeb (2017, p. 172), B_A deve ser entendido como o conjunto de "*A*–mundos" mais plausíveis, ou seja, o conjunto de mundos mais plausível dentre aqueles que satisfazem A. Desse modo, previsivelmente, B_W é o conjunto de mundos que são mais plausíveis no geral.

Com tais elementos em mãos, agora já somos capazes de determinar um processo de revisão qualitativa em termos de $TH_{\mathcal{Y}}$, ao estilo AGM. Seja $Bel(\cdot \mid \cdot)$ uma crença condicional determinada por $\mathbb{S}$. Como já foi dito aqui, uma revisão de crença em um sistema de esferas é determinada por uma função de seleção da seguinte maneira: para qualquer $A, Y \in \mathcal{A}$, $Bel(A \mid Y)$ *sse* $A \supseteq Y \cap S_Y$, onde S_Y é a menor efera que intersecciona Y[27]. Como $Y \cap S_Y$ é precisamente o conjunto dos mundos mais plausíveis que satisfazem Y, isto é, B_Y, temos que $Bel(A \mid Y)$ *sse* $A \supseteq B_Y$. Em outras palavras, o agente acredita em A condicional sobre Y se, e somente se, todos os $Y-$ *mundos* mais plausíveis são $A-$ *mundos*. E assim $TH_{\mathcal{Y}}$ emerge de um sistema de esferas aplicado a núcleos de crença.

E é aí precisamente que $TH_{\mathcal{Y}}$ coopera com a semântica de esferas na teoria AGM. Segundo Leitgeb (2017, p. 170), a estrutura semântica de esferas na teoria AGM é simplesmente um conjunto de escolhas possíveis, mas não atualizadas, para os candidatos ao núcleo B_W, onde tais escolhas são realmente mais prudentes do que a proposição B_W na questão que gera o conjunto de crenças atual. O aninhamento de esferas proporciona modos de ser para crenças do agente, caso ele fosse mais prudente. Como vimos brevemente acima, e nos aprofundaremos logo mais, um dos resultados mais significativos da teoria de Leitgeb garante que a classe de conjuntos *P*–estáveis de probabilidade menor que 1 é ordenado pela relação de subconjunto do mesmo modo que o sistema de esferas de Groove para os operadores de revisão da teoria AGM: a noção de estabilidade demanda a estrutura formal de um sistema de esferas ou de um pre-ordenamento total de mundos (LEITGEB, 2017, p. 171).

Agora avançando um pouco mais, para construir uma versão diacrônica de $TH_{\mathcal{Y}}$, Leitgeb adiciona aos postulados sincrônicos iniciais, conforme apresentados na Seção 1.2.1, mais dois postulados específicos para revisão. O primeiro deles é restrito aos casos de revisão (ou melhor, uma expansão no sentido AGM) onde a evidência adquirida não contradiz o conjunto inicial de crença (LEITGEB, 2017, p. 184):

- Para todo $Y \in \mathcal{A}$ tal que $Y \cap B_W \neq \emptyset$, e para todo $A \in \mathcal{A}$:
  $Bel(A \mid Y)$ se, e somente se, $A \supseteq Y \cap B_W$

[27] Recomendo ao leitor que retome a Seção 4.2.1 caso sinta necessidade.

Contudo, o caso que realmente nos interessa é aquele onde temos uma revisão genuína (no sentido AGM do termo), e a evidência adquirida é incompatível com [**B**]. Nesse caso, tal norma consiste na união dos axiomas $\mathbf{B}^{*}7$ e $\mathbf{B}^{*}8$ da teoria AGM, e chamaremos aqui de *MR*, em referencial ao princípio de Monotonicidade Racional. Nesse caso, não exigimos mais que $Y \cap B_W \neq \emptyset$, ao invés disso, Y irá variar conforme a proposição sob a qual estamos condicionalizando. Mais precisamente (LEITGEB, 2017, p. 215):

(MR) Para todo $E, Y \in \mathcal{A}$ tal que $Y \cap B_E \neq \emptyset$, e para todo $A \in \mathcal{A}$:
$Bel(A \mid E \cap Y)$ se, e somente se, $A \supseteq Y \cap B_E$

onde B_E é o menor conjunto de tudo que o agente acredita condicional à E, ou seja, o conjunto $\mathbf{B}^{*}_{E}$, ou em uma notação alternativa, $Bel(\cdot \mid E)$, o que implica que $B_E \subseteq E$. E isso é exatamente a noção de núcleo de crença B_W recorrentemente mencionada aqui.

O que (*MR*) afirma, em outras palavras, é que se a proposição Y é consistente com B_E, isto e, se Y é consistente com tudo que o agente acredita condicionalmente sobre E, então ele acredita em A condicional sobre a intersecção $Y \cap E$ somente no caso em que A é uma consequência lógica de $Y \cap B_E$. Podemos notar que (*MR*) é simplesmente uma versão do postulado de Monotonicidade Racional na lógica não-monotônica. Assim como na teoria AGM, (*MR*), em conjunto com os postulados iniciais para crença, não determina um único conjunto de crenças. O que a teoria de Leitgeb (2017, p. 216) adiciona na teoria AGM original é que agora a distribuição de probabilidades subjetivas P irá atuar em conjunto: Alguns conjuntos de crença [**B**] irão excluir certas distribuições P, e vice versa.

E aqui uma das diferenças fundamentais entre $TH_{\mathcal{Y}}$ e *PL* surge na forma como, principalmente, (*MR*) aparece na teoria qualitativa de revisão de crença: enquanto *PL* assume um modelo Preferencial que rejeita Monotonicidade Racional, $TH_{\mathcal{Y}}$ assume a teoria AGM onde (*MR*) é axioma. Como consequência, a ordem de plausibilidade entre mundos possíveis no primeiro caso é parcial: existem casos onde w_i e w_j não se relacionam por meio de um ordenamento $\preccurlyeq$[28]. Já no segundo caso, a ordem de plausibilidade é total: para quaisquer w_i e w_j, ou $w_i \preccurlyeq w_j$, ou $w_j \preccurlyeq w_i$. Nesse último caso, também dizemos que se trata de um *ranqueamento* de mundos possíveis. E é exatamente o postulado (*MR*) o principal resposável por isso, ele expressa a totalidade, ou linearidade, da pré-ordem entre mundos[29]. Como consequência, a semântica de

[28] Contudo, se o limiar t em *PL* fosse estabelecido como sendo igual a 1 (o que é uma possibilidade rejeitada por Lin e Kelly), [**B**] seria determinado a partir de uma pré-ordem total, assim como no sistema de esferas da teoria AGM.

[29] Como vimos na Seção 4.2.2, modelos que satisfazem os requisitos de um sistema de esferas para inferências não-monotônicas $\mid\!\sim$ são usualmente chamados de “modelos ranqueados” onde, a relação de plausibilidade $\prec$ entre mundos é parcial, e a relação de consequência $\mid\!\sim$ satisfaz Monotonicidade Racional, formando assim um sistema **R**.

revisão de $TH_{\mathcal{Y}}$ correspondente ao aninhamento de esferas.

Segundo Leitgeb (2017, p. 217), a exigência de totalidade possui três benefícios para a presente discussão: (i) reforça simplicidade, o que cai bem com a ideia de que crenças (qualitativas) são mais simples em relação a probabilidades subjetivas; (ii) condiz com a linearidade do lado numérico da probabilidade subjetiva; e mais importante, (iii) se alguém mantém a distribuição P fixa, os diferentes candidatos ao conjunto B_W serão P-estáveis (conforme veremos com mais detalhes no teorema abaixo). Para Leitgeb (2017, p. 187), (MR) se justifica como requisito para crença condicional simplesmente pelo fato de que, sem ele, crença qualitativa não teria sua capacidade de simplificação, o que é essencial para ela. Nesse sentido, para o autor, crença e probabilidade subjetiva desempenham papéis doxástico distintos, e a crítica de Lin e Kelly simplesmente mistura as condições sobre estados doxásticos qualitativos e quantitativos.

No caso do contraexemplo estilo Gettier oferecido como um argumento contra Mononicidade Racional, como vimos na seção anterior, Leitgeb (2017, p. 187) argumenta que, se a crença inicial relevante é "Nogot ou Havit tem um Ford", então cada uma das opções deve ser suficientemente mais provável do que a opção "Nenhum deles tem um Ford". E a explicação para tal condição se fundamenta em um argumento de estabilidade: se isso não for o caso, a crença "Nogot ou Havit" não seria estável sob condicionalização sobre o que o agente toma como uma possibilidade real. Alguns filósofos da ciência[30] referem-se a tal propriedade como um princípio de "tenacidade": uma hipótese verdadeira, uma vez aceita, não pode ser rejeitada em um momento seguinte baseada em uma nova evidência verdadeira.

Finalmente, reproduzo agora o principal resultado formal desta seção, onde boa parte das afirmações feitas até aqui são apropriadamente justificadas. O teorema a seguir se utiliza de um postulado que afirma que ter probabilidade subjetiva condicional acima de 1/2 é uma condição necessária para ter crença condicional. Isso corresponde a versão esquerda-para-direita de TL, e Lin e Kelly (2021, p. 199) a chamam de "Princípio de Alta Probabilidade".

Teorema 4.3.3 (Teorema da Representação para Crença Condicional)**.** *Seja* [***B***] *o conjunto de crenças do agente, e P uma distribuição de probabilidade sobre os elementos de W. As duas afirmações a seguir são equivalentes:*

*(**I**) P satisfaz os postulados iniciais para probabilidade subjetiva,* [***B***] *satisfaz os postulados iniciais para crença, mais MR, e ambos P e* [***B***] *satisfazem um "Princípio de Alta Probabilidade" (PAP):*

$$\textit{Para todo } Y > 0\textit{, e todo } A \in \mathcal{A}\textit{, se } Bel(A \mid Y) \textit{ então } P(A \mid Y) > 1/2.$$

[30] Leitgeb faz referência a Hilary Putnam.

(***II***) *P satisfaz os postulados iniciais para probabilidade subjetiva. P e $\mathcal{A}$ são tais que $\mathcal{A}$ contém um menor conjunto de probabilidade 1, e existe uma classe $\mathbb{S}$ de proposições não-vazias e P–estáveis em $\mathcal{A}$ tal que: (i) $\mathbb{S}$ contém o menor conjunto de probabilidade 1 em $\mathcal{A}$, e (ii) todos outros membros de $\mathbb{S}$ possuem probabilidade menor do que 1. Além disso:*

– *Para todo $Y \in \mathcal{A}$ com $P(Y) > 0$: se S_1 é o menor membro de $\mathbb{S}$ no que diz respeito a relação de subconjunto, e S_1 é o menor membro de $\mathbb{S}$ para o qual temos $Y \cap S_1 \neq \emptyset$. Então, para todo $A \in \mathcal{A}$:*

$$Bel(A \mid Y) \text{ sse } A \supseteq Y \cap S_1$$

.

A demonstração da ida, isto é, I para II, é bastante sofisticada e o leitor interessado pode conferí-la no Apêndice B, ou no próprio texto do autor, em Leitgeb (2017, p. 220). Já a volta, de II para I, é menos complexa e nos é interessante no seguinte sentido: sua primeira parte nos oferece uma receita para a construção de modelos para a união de todos os postulados para [**B**] a partir de qualquer distribuição *P* da qual exista um menor conjunto de probabilidade 1. Resumidamente, o procedimento é o seguinte: Tome algum conjunto *P*–estável de probabilidade menor do que 1 (se existir), e tome também o menor conjunto de probabilidade 1. Coloque-os juntos em $\mathbb{S}$. Então defina crença condicional tal como aparece na parte II do teorema, e temos assim o modelo esperado.

Agora, a partir do procedimento acima, podemos demonstrar que a noção de crença condicional em II satisfaz os postulados para um PP diacrônico mencionados em I. Primeiramente, *P* satisfaz os postulados iniciais para probabilidade subjetiva por hipótese. Se $S_1 \in \mathbb{S}$ é um núcleo de crença e, por hipótese, é *P*–estável, pelo Teorema da Representação 3.1.1 para $TH_{\mathcal{Y}}$, apresentado no capítulo 3, temos que S_1 satisfaz os postulados iniciais para crença. Agora vamos assumir que o antecedente de *PAP* é satisfeito, ou seja, $Bel(A \mid Y)$. Por (II) temos que $A \supseteq S_1 \cap Y$, logo, $A \cap Y \supseteq S_1 \cap Y$, o que implica, pelas propriedades da teoria da probabilidade, que $P(A \cap Y) \geq P(S_1 \cap Y)$. Como S_1 é *P*–estável, e $S_1 \cap Y \neq \emptyset$, temos que $P(S_1 \mid Y) > 1/2$, pela Definição 22 de *P*–estabilidade. Tomando esses dois fatos em conjunto, podemos afirmar que $P(A \mid Y) > 1/2$. Finalmente, basta mostrar que a noção relevante de crença condicional também satisfaz (*MR*), o que aparece de modo análogo na demonstração de Groove para o sistema de esferas (que também pode ser conferida no Apêndice B).

Isso nos mostra que todas esferas em um sistema de revisão AGM são *P*-estáveis. O processo de revisão de crença da teoria AGM por meio de um sistema de esferas diz como o agente deve operar quando confrontado com uma evidência que contradiz seu conjunto atual de crenças. Isso corresponde basicamente ao conjunto de crenças condicionais do agente. De acordo com o teorema acima, tais esferas são conjuntos P–estáveis, ou seja, são conjuntos que são candidatos possíveis para o núcleo B_W. Segundo Leitgeb (2017, p. 223), a estrutura de esferas na semântica da teoria AGM é precisamente um conjunto de escolhas possíveis, mas não atualizadas, para o núcleo B_W.

Para finalizar, antes de encerrar este tópico é importante falar um pouco mais sobre como a divergência acima sobre a aceitação ou não da Monotonicidade Racional, como norma para revisão de crença, impacta a discussão sobre o modo como cada teoria se posiciona no debate sobre a natureza e racionalidade de um Princípio Ponte. Como já mencionei algumas vezes, ao insistir na estabilidade como uma norma da crença, Leitgeb abre mão da possibilidade de reduzir a racionalidade do conjunto de crenças à racionalidade do conjunto de probabilidades subjetivas. Segundo Leitgeb (2017, p. 230), uma possibilidade de redução teria a forma de "crença = probabilidade subjetiva + x", onde x pode ser definido como algum elemento prático, como contexto, preferências, ou inclinações do agente. Mas a racionalidade da crença não pode ser determinada simplesmente em função do conjunto de probabilidades subjetivas.

Como vimos pelo teorema acima, dada uma distribuição de probabilidade P, é natural que exista mais do que apenas um conjunto de crenças $[\mathbf{B}]$ que satisfaz os postulados AGM, os postulados probabilísticos e lógicos apresentados no Capítulo 1. Cada um desses conjuntos possíveis de crença corresponde a um sistema de esferas contendo conjuntos P–estáveis. E podem haver inúmeros desses sistemas satisfazendo as condições desejadas. Ou seja, $TH_{\mathcal{Y}}$ como PP, tanto para crença condicional como não-condicional, pode levar a mesma distribuição P em dois conjuntos de crenças distintos $[\mathbf{B}]$ e $[\mathbf{B}']$. E isso impossibilita uma redução na forma de uma definibilidade de $[\mathbf{B}]$ por meio de uma função de aceitação (ou seja, um PP) que possua como argumento uma medida de probabilidade qualquer.

Em uma de suas primeiras formulações da Tese Humeana (ainda chamada de Teoria da Estabilidade), Leitgeb (2013) propõe tal tipo de redução assumindo que o conjunto $[\mathbf{B}]$ final, determinado por meio de um núcleo B_W, deveria ser aquele que "maximizava o conjunto de crenças do agente", ou escolhendo o menor conjunto P–estável dentre as opções disponíveis. Todavia, pouco tempo depois o autor abandonou o projeto reducionista, uma vez que, para ele, não parece haver nada de mais "racional" em selecionar um conjunto de crença mais prudente do que um mais ousado (LEITGEB, 2017, p. 234).

Mais do que um projeto de redução, a proposta de Leitgeb é oferecer uma norma de coerência na forma de um Princípio Ponte, o que chamamos aqui, desde o início, de um requisito racional estrutural. E como requisito racional diacrônico, $TH_{\mathcal{Y}}$ funciona de modo bastante eficiente: condicionalização bayesiana e mudança de visão na forma de revisão AGM funcionam em plena harmonia, de tal modo que, se P e $[\mathbf{B}]$ satisfazem

$TH_{\mathcal{Y}}$ em t_1, então $P(\cdot \mid E)$ e $Bel(\cdot \mid E)$ também satisfazem quando atualizadas em t_2.

Por outro lado, o que *PL* exige é mais forte: a teoria assume que *P* e [**B**] comutam entre si no que diz respeito a um PP diacrônico. Isso significa que o PP relevante é uma função *f* de *P* para [**B**], isto é, $f(P) = [\mathbf{B}]_P$. E do lado diacrônico temos que $f(P(A \mid E)) = Bel(f(P) \mid E)$. Ou seja, precisamente a relevante noção de rastreio apresentada anteriormente: não importa se atualizamos primeiro probabilisticamente e só então determinamos o conjunto de crenças por meio de *f*; ou se determinamos a crença primeiro por *f* e então revisamos a crença condicionalmente. Isso proporciona a possibilidade de um argumento de *sobreveniência* entre tais estados: qualquer mudança no conjunto de crenças (qualitativas) representa uma mudança no conjunto probabilidades subjetivas. Em outras palavras, sobreveniência é entendida como uma função de *P* para [**B**], e serve perfeitamente a um projeto reducionista.

Já do lado humeano da discussão, *P* e [**B**] não comutam entre si: embora condicionalização bayesiana preserve estabilidade de conjuntos de crenças anteriores, ela pode gerar novos conjuntos estáveis que não correspondem a revisão qualitativa. Isso ocorre, como vimos, fundamentalmente devido ao fato de existirem mais de uma opção racional dentre as opções de núcleos B_W disponíveis. Será então que podemos chegar em alguma conclusão sobre qual das teorias é mais eficiente em propor um PP diacrônico como um requisito racional?

4.4 POR QUE NÃO AMBOS?

Para finalizar o presente trabalho, pretendo lançar luz sobre o desafio no qual encerramos o tópico anterior. Como vimos durante todo o capítulo, nosso desafio em torno da busca por um PP diacrônico apropriado enfrenta um profundo conflito sobre qual o modelo formal mais adequado para representar uma revisão qualitativa de crença, ou melhor, uma mudança de visão racional. Vimos que uma das vantagens de *PL* é o fato de que, ao utilizar uma lógica não-monotônica e uma semântica preferencial na revisão de crenças, a teoria é capaz de oferecer um espelhamento perfeito com a condicionalização bayesiana, condição que Lin e Kelly chamam de *rastreamento*. Mas isso seria um ganho significativo se nosso objetivo fosse oferecer uma proposta de PP que reduzisse um estado doxástico à outro, o que não é o caso. Como foi dito desde o primeiro capítulo, nossa proposta é encontrar um PP na forma de um requisito racional estrutural, e nesse sentido, tanto *PL* quanto $TH_{\mathcal{Y}}$ satisfazem bem tal condição, tanto em suas abordagens estáticas quanto dinâmicas.

Na medida em que *PL* é capaz de oferecer algo que $TH_{\mathcal{Y}}$ não é capaz, isto é, rastreabilidade (e consideramos isso algo positivo), precisamos olhar para o que $TH_{\mathcal{Y}}$ oferece que *PL* não é capaz de ofertar. Como vimos, a resposta para isso é a condição de *estabilidade*. Mas por que isso importa? Como vimos no Capítulo 3, parte da explicação repousa sobre uma possível interpretação dos escritos de Hume sobre a natureza da crença. Mas mais do que isso, não é difícil encontrar referências em defesa da crença como um estado estável, principalmente relacionado ao debate sobre racionalidade prática. Para alguns autores, estabilidade parece importante para crença paralelamente ao modo como estabilidade é importante para a noção de “intenção”. Nas palavras de Richard Holton (2014, p. 33), “minha habilidade para operar no mundo depende de que eu mantenha muitas questões fechadas”. No que diz respeito à racionalidade prática, a tese mais provável é que a estabilidade da intenção seja derivada justamente da estabilidade da crença[31]. Mas não precisamos nos aprofundar demasiadamente nessa discussão aqui.

O ponto aqui é que estabilidade parece importar, pelo menos para um conceito relevante de intenção. Um passo seguinte seria investigar se a noção de estabilidade em Leitgeb é a melhor disponível para uma noção intuitiva de “crenças e intenções estáveis”. Como vimos no Capítulo 3, em $TH_{\mathcal{Y}}$, “estabilidade” é entendida como “estável sob condicionalização probabilística”: quando condicionalizada por uma crença que o agente considera uma possibilidade epistêmica, a crença inicial não é abandonada. Tal definição parece compatível com a noção de estabilidade defendida pelos teóricos da racionalidade prática:

> Podemos procurar indefinidamente por evidências para nossa proposição, mas precisamos saber quando interromper a busca e equipar nossa conclusão com um razoável grau de estabilidade, de tal modo que possamos raciocinar e planejar com base nisso. Aqui, novamente, precisamos encontrar um balanço entre sensibilidade ao mundo e estabilidade. (HOLTON, 2014, p. 29)

E o que Leitgeb oferece é exatamente uma definição precisa disso que Holton chama de “sensibilidade ao mundo”. E isso fica ainda mais claro na seguinte passagem:

> Estabilidade requer que o limiar no qual uma nova evidência conduzirá a uma reconsideração [...] seja maior do que o limiar do qual precisou ultrapassar para ser considerada em sua informação inicial. O quão alto deve ser dependerá de diversos fatores. (HOLTON, 2014, p. 15)

Onde tudo isso nos leva? Parece que estabilidade não é algo simplesmente descartável, sem qualquer prejuízo, e $TH_{\mathcal{Y}}$ parece cumprir bem tal demanda. Contudo, como vimos no tópico anterior, estabilidade é incompatível com rastreamento bayesiano. Feitas as devidas considerações, minha proposta aqui é que, apesar de incompatíveis, não precisamos, necessariamente, decidir entre uma delas e descartar a outra como inapropriada. Isso é possível uma vez que não estamos disputando aqui qual é “ A teoria verdadeira” sobre nossas atitudes doxásticas, e a ponte racional que pode ser construída entre elas. O que estamos fazendo aqui é tentando modelar formalmente alguns fenômenos que consideramos filosoficamente interessantes. Conforme

[31] Em Rui (2018) eu abordo com mais detalhes a relação entre os conceitos de crença, intenção, e racionalidade prática. Mais sobre o assunto também pode ser encontrado em Bratman (1987) e Broome (2013).

argumento no Apêndice A, uma boa teoria da racionalidade e do raciocínio pode se utilizar do modo como nós “de fato” raciocinamos, de nossas intuições normativas sobre o que é considerado “mais racional”, e também pode se utilizar do trabalho de modelagem formal que é capaz de levar em conta todos esses elementos, simultaneamente, e nos mostrar as consequências desejáveis, e indesejáveis, da nossa teoria.

Nesse sentido, podemos separar as duas opções de PP diacrônicos como possibilidades de modelos formais para uma teoria do raciocínio, ou mudança de visão: se nosso objetivo for trabalhar com casos onde o que nos interessa é uma correspondência perfeita entre crenças qualitativas e probabilidades subjetivas, *PL* é claramente a melhor opção para um PP diacrônico; mas se o objetivo for lidar com situações em que crenças precisam servir como ferramentas que simplifiquem nosso raciocínio e tomada de decisão, pode ser que a melhor estratégia seja utilizar um modelo com crenças estáveis, e $TH_{\mathcal{Y}}$ é obviamente a melhor candidata. Resumindo, ambas podem conviver pacificamente, cada uma sendo a ferramenta para o teórico que visa modelar um fenômeno particular, mesmo que não possam ser utilizadas simultaneamente, no mesmo projeto.

Além disso, acredito que a discussão sobre um PP dinâmico, sobre teorias da mudança de visão racional e condicionalização bayesiana, possa ser de maior relevância filosófica do que a discussão tradicional sobre a consistência e fechamento lógico em um PP estático. Pode ser que os paradoxos de consistência, da loteria e do prefácio, apenas nos conduziram para um problema ainda mais profundo sobre teorias qualitativas e quantitativas do raciocínio. Nesse sentido, defendo que a existência de modelos incompatíveis seja apenas uma ampliação de horizontes sobre como devemos modelar tal fenômeno relevante, isto é, nossa mudança de visão racional: talvez seja necessário fazer isso de um modo mais semelhante ao que humanos fazem com suas crenças qualitativas e em suas tomadas de decisão, o que implica em um modelo com estabilidade; mas também pode ser que seja necessário trabalhar com um modelo mais idealizado que rastreia condicionalização bayesiana.

Por fim, quero reforçar o fato, defendido no Apêndice A, de que apenas fomos capazes de chegar em questões tão técnicas e profundas sobre raciocínio, ou mudança de visão, graças a estratégia metodológica utilizada desde o início da Tese: o trabalho de modelagem formal como ferramenta para analisar problemas epistêmicos. Se não tivéssemos adotado tal estratégia metodológica frente aos pertinentes problemas causados pelos paradoxos da loteria e do prefácio, dificilmente haveria a possibilidade de qualquer avanço em direção ao que vimos no presente capítulo. E como dito no parágrafo acima, me parece que compreender as minúcias do problema da dinâmica do raciocínio com estados qualitativos e quantitativos seja uma tarefa ainda mais rica do que a mera solução de um problema de consistência gerado inicialmente pelos paradoxos. Foi só a partir da construção de modelos formais que fomos capazes de chegar até aqui, o que considero algo digno de certo mérito.

CONCLUSÃO

Partindo da suposição de que nosso estudo sobre a racionalidade de estados doxásticos deve lidar com concepções qualitativas e quantitativas de crenças, a presente Tese se propôs a examinar cuidadosamente as principais possibilidade teóricas de compreender como esses distintos estados se relacionam. Considero que este trabalho contribuiu para a literatura sobre o tema na medida em que (i) sistematizou uma discussão ainda em processo de construção; (ii) aproximou teoricamente correntes ainda um tanto quanto distantes, como é o caso das discussões sobre a natureza da racionalidade e sobre epistemologia formal; e (iii) propôs novas alternativas de respostas para problemas pontuais ainda em discussão pela literatura relevante que investiga o tema.

Ao longo do primeiro capítulo argumentei que existe uma concepção de "racionalidade estrutural" que se encaixa muito bem com os debates mais avançados em epistemologia formal, mais precisamente, o debate sobre a racionalidade da crença. Com as devidas ressalvas mencionadas durante o capítulo, essa é uma estratégia muito pouco utilizada quando queremos discutir modelos formais quantitativos e qualitativos para crença. Ainda, um passo importante foi estabelecer que nossa busca por um Princípio Ponte (PP) que conecte a racionalidade da crença e da probabilidade subjetiva pode ser entendido em termos de um requisito racional estrutural, isto é, uma norma de escopo amplo e global, tal como definido na Seção 1.1.1. Isso significa que, graças a condição de escopo amplo, um PP entendido como um requisito racional não precisa ser uma proposta de redução entre os nossos estados doxásticos relevantes, mas simplesmente uma regra da racionalidade pontificando dois reinos distintos: o qualitativo e o quantitativo.

Como consequência da concepção de racionalidade mencionada no parágrafo acima, encerramos o Capítulo 2 argumentando que um PP na forma de uma Teoria da Decisão Cognitiva não seria nosso alvo primário de investigação, uma vez que a teoria não é capaz de fornecer um PP ao estilo de um requisito racional estrutural, uma vez que processos decisórios são claramente assimétricos. Ainda sobre a relevância da noção estrutural de racionalidade para nossa investigação, terminamos Capítulo 4 com o grande momento da Tese onde tentamos responder qual PP considerávamos mais adequado, após mais de cem páginas de investigação e análise das teorias relevantes. Como vimos, na medida em que o propósito de nosso PP é ser um requisito racional estrutural, a condição de rastreabilidade (um dos grandes méritos) da Teoria Probalógica (*PL*) não oferece uma razão adicional em favor de tal PP, uma vez que um requisito racional exige simplesmente uma norma de coerência, não um princípio de redução ou sobreveniência. Ou seja, a estratégia de entender um PP como um tipo de requisito racional estrutural joga a favor da Tese Humeana ($TH_{\mathcal{Y}}$) de Leitgeb, uma vez que não precisamos de um mapeamento biunívoco entre os estados doxásticos, o que sua teoria é incapaz de oferecer.

No que diz respeito ao conflito entre *PL* e $TH_{\mathcal{Y}}$, encerramos o último capítulo mostrando que, apesar de incompatíveis, podemos adotar uma perspectiva moderada sobre nosso propósito e aceitar que, principalmente de um ponto de vista dinâmico, ambas podem ser mantidas como opções válidas de PP, e o que vai decidir sobre qual deverá ser utilizada é o objetivo particular do teórico interessado em modelar um específico fenômeno epistêmico. Contudo, é importante mencionar aqui que, ainda de um ponto de vista puramente estático, concluímos o terceiro capítulo apresentado uma dupla de objeções à teoria de Leitgeb. Recordando brevemente, vimos que a referida teoria acaba sendo demasiadamente cética com limiares bem acima de 1/2, e com limiares mais baixos, próximos de 1/2, ela sofre por ser "mente fechada", isto é, o agente irá considerar sempre um conjunto bastante restrito de possibilidades doxásticas. Nesse sentido, Acredito que *PL* possui uma ligeira vantagem teórica em relação a $TH_{\mathcal{Y}}$, uma vez que é mais eficiente em sua capacidade de entregar um conjunto informativo (não cético) de crenças.

É claro que a discussão aqui presente ainda está processo de consolidação, e acredito que, em alguns anos, boa parte do que foi apresentado poderá se mostrar desatualizado ou até mesmo fora de propósito. Mas isso faz parte da própria essência do trabalho filosófico e não deve tirar o mérito do que está sendo entregue no presente momento. E, apesar da considerável extensão deste trabalho, é preciso admitir aqui que muitas questões relevantes acabaram ficando em aberto ou não tiveram, possivelmente, a devida atenção recebida. Entre elas eu colocaria (i) a relevância de uma abordagem mais completa sobre lógica epistêmica com operadores para crença e conhecimento e suas relações de segunda ordem; (ii) dentro do segundo capítulo, uma noção mais sofisticada de informação na determinação da utilidade epistêmica de uma crença; e (iii) um aprofundamento maior sobre a diferença entre o paradoxo da loteria e o paradoxo do prefácio, e quais as implicações disso para cada PP em questão, uma vez que ambos foram tratados, por uma escolha pessoal, como apontando o mesmo conjunto de problemas na busca por um Princípio Ponte.

REFERÊNCIAS

ALCHOURRÓN, Carlos E; GÄRDENFORS, Peter; MAKINSON, David. On the logic of theory change: Partial meet contraction and revision functions. **Journal of symbolic logic**, JSTOR, p. 510–530, 1985.

ARLÓ-COSTA, Horacio. Qualitative and probabilistic models of full belief. *In*: BUSS; R.; HÁJEK, P.; PUDLÁK, P. (Ed.). **Proceedings of logic colloquium**. [*S.l.*: *s.n.*], 1999. Cambridge University Press, p. 25–43.

ARLÓ-COSTA, Horacio; PEDERSEN, Arthur Paul. Belief and probability: A general theory of probability cores. **International Journal of Approximate Reasoning**, Elsevier, v. 53, n. 3, p. 293–315, 2012.

ARNTZENIUS, Frank. Some problems for conditionalization and reflection. **The Journal of Philosophy**, v. 100, n. 7, p. 356–370, 2003.

BINMORE, Ken. Rationality and backward induction. **Journal of Economic Methodology**, Taylor & Francis, v. 4, n. 1, p. 23–41, 1997.

BRADLEY, Darren. **A critical introduction to formal epistemology**. New York, USA e London, UK: Bloomsbury Publishing, 2015.

BRATMAN, Michael. **Intention, plans, and practical reason**. Cambridge, MA: Harvard University Press, 1987. v. 10.

BRIGGS, Rachael. Distorted reflection. **Philosophical Review**, Duke University Press, v. 118, n. 1, p. 59–85, 2009.

BROOME, John. Normative requirements. **Ratio**, Wiley Online Library, v. 12, n. 4, p. 398–419, 1999.

BROOME, John. **Rationality Through Reasoning**. Chichester: John Wiley & Sons, 2013.

BRUNERO, John. The scope of rational requirements. **The Philosophical Quarterly**, Blackwell Publishers Ltd, Oxford UK e Boston, USA, v. 60, n. 238, p. 28–49, 2010.

CARNAP, Rudolf. **Logical Foundations of Probability**. Chicago: The University of Chicago Press, 1950.

CHAN, Timothy. **The aim of belief**. New York: Oxford University Press, 2013.

CHRISTENSEN, David. Clever bookies and coherent beliefs. **The Philosophical Review**, JSTOR, v. 100, n. 2, p. 229–247, 1991.

CHRISTENSEN, David. **Putting logic in its place: Formal constraints on rational belief**. New York: Oxford University Press, 2004.

COLYVAN, Mark. Idealisations in normative models. **Synthese**, Springer, v. 190, n. 8, p. 1337–1350, 2013.

CORTZEN, Allan; WEISSTEIN, Eric W. **Measure From MathWorld–A Wolfram Web Resource**. Disponível em: `https://mathworld.wolfram.com/Measure.html`. accessed: 19.06.2022).

DEROSE, Keith. Contextualism and knowledge attributions. **Philosophy and phenomenological research**, v. 52, n. 4, p. 913–929, 1992.

DORST, Kevin. Lockeans maximize expected accuracy. **Mind**, Oxford University Press, v. 128, n. 509, p. 175–211, 2017.

DOUVEN, Igor; ROTT, Hans. From probabilities to categorical beliefs: Going beyond toy models. **Journal of Logic and Computation**, Oxford University Press, v. 28, n. 6, p. 1099–1124, 2018.

EASWARAN, Kenny. Dr. Truthlove or: How I learned to stop worrying and love Bayesian probabilities. **Noûs**, Wiley Online Library, v. 50, n. 4, p. 816–853, 2016.

EASWARAN, Kenny; FITELSON, Branden. Accuracy, coherence, and evidence. *In*: TAMAR GENDLER, John Hawthorne (Ed.). **Oxford studies in epistemology**. [*S.l.*]: Oxford University Press Oxford, 2015. v. 5. P. 61–96.

ENGEL, Pascal. In defence of normativism about the aim of belief. *In*: CHAN, Timothy; CHAN, Timothy Hoo Wai (Ed.). **The aim of belief**. New York: Oxford University Press, 2013. P. 32–63.

FINETTI, Bruno de. Methods for discriminating levels of partial knowledge concerning a test item. **British Journal of Mathematical and Statistical Psychology**, Wiley Online Library, v. 18, n. 1, p. 87–123, 1965.

FITELSON, Branden; EASWARAN, Kenny; MCCARTHY, David. **forthcoming Coherence**. [*S.l.*: *s.n.*]. Disponível em:
`http://fitelson.org/coherence/coherence_duke.pdf`. accessed: 29.03.2021).

FOLEY, Richard. The Epistemology of Belief and the Epistemology of Degrees of Belief. **American Philosophical Quaterly**, v. 29, n. 0, p. 111–121, 1992.

FRIEDMAN, Milton. **Essays in positive economics**. Chicago: University of Chicago press, 1953.

GABBAY, Dov M; HOGGER, Christopher John; ROBINSON, John Alan. **Handbook of logic in artificial intelligence and logic programming: Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning**. Oxford: Clarendon Press, 1994.

GÄRDENFORS, Peter. **Knowledge in flux: Modeling the dynamics of epistemic states.** Massachusetts: The MIT press, 1988.

GENIN, Konstantin. Full & partial belief. *In*:
PETTIGREW RICHARD E WEISBERG, Jonathan (Ed.). **The Open Handbook of Formal Epistemology**. [*S.l.*]: Published Open Access by PhilPapers, 2019.
P. 437–498.

GENIN, Konstantin; HUBER, Franz. Formal Representations of Belief. *In*:
ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Winter 2020.
[*S.l.*]: Metaphysics Research Lab, Stanford University, 2020.

GIBBARD, Allan. Rational credence and the value of truth. **Oxford studies in epistemology**, Oxford University Press Oxford, v. 2, p. 143–164, 2007.

GIBBARD, Allan. Truth and correct belief. **Philosophical Issues**, JSTOR, v. 15, p. 338–350, 2005.

GIBBARD, Allan. **Wise choices, apt feelings: A theory of normative judgment**.
[*S.l.*]: Harvard University Press, 1990.

GROVE, Adam. Two modellings for theory change. **Journal of philosophical logic**, JSTOR, p. 157–170, 1988.

HÁJEK, Alan. Pascal's Wager. *In*: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Summer 2018. [*S.l.*]: Metaphysics Research Lab, Stanford University, 2018.

HANSSON, Sven Ove. Belief Change. *In*: HANSSON, Sven Ove; HENDRICKS, Vincent (Ed.). **Introduction to Formal Philosophy**. Spring. Cham, Switzerland: Springer, 2018.

HANSSON, Sven Ove. Logic of Belief Revision. *In*: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Spring 2022. [*S.l.*]: Metaphysics Research Lab, Stanford University, 2022.

HARMAN, Gilbert. **Change in view: Principles of reasoning.** [*S.l.*]: The MIT Press, 1986.

HAWTHORNE, James. The Lockean thesis and the logic of belief. *In*: DEGREES of belief. [*S.l.*]: Springer, 2009. P. 49–74.

HEDDEN, Brian. Time-slice rationality. **Mind**, Oxford University Press, v. 124, n. 494, p. 449–491, 2015.

HEMPEL, Carl G. Deductive-Nomological vs. Statistical Explanation. *In*: MAXWELL, H. Fiegl & G. (Ed.). **Studies in the Philosophy of Science**. Minnesota: University of Minnesota Press, 1962. P. 98–169.

HINTIKKA, Jaakko. **Knowledge and Belief: An Introduction to the Logic of the Two Notions**. New York: Cornell University Press, 1962.

HOLTON, Richard. Intention as a Model for Belief. *In*: VARGAS, Manuel; YAFFE, Gideon (Ed.). **Rational and Social Agency: The Philosophy of Michael Bratman**. New York: Oxford University Press, 2014. cap. 2, p. 12–37.

HOOKER, Brad; STREUMER, Bart. Procedural and substantive practical rationality. *In*: THE Oxford handbook of rationality. [*S.l.*]: Oxford University Press, 2004. P. 57–74.

HUBER, Franz. Belief and degrees of belief. *In*: DEGREES of belief. [*S.l.*]: Springer, 2009. P. 1–33.

JAMES, William. **The will to believe: And other essays in popular philosophy, and Human immortality**. [*S.l.*]: Courier Corporation, 1956. v. 291.

JEFFREY, Richard. Dracula Meets Wolfman: Acceptance vs. Partial Belief. *In*: SWAIN, Marshall (Ed.). **Induction, Acceptance, and Rational Belief**. Dordrecht: editora, 1970. cap. x, p. 157–185.

JEFFREY, Richard. **The logic of decision**. [*S.l.*]: University of Chicago Press, 1990.

JEFFREY, Richard. Valuation and acceptance of scientific hypotheses. **Philosophy of Science**, Williams e Wilkins Co., v. 23, n. 3, p. 237–246, 1956.

JOYCE, James M. A nonpragmatic vindication of probabilism. **Philosophy of science**, University of Chicago Press, v. 65, n. 4, p. 575–603, 1998.

JOYCE, James M. How probabilities reflect evidence. **Philosophical perspectives**, JSTOR, v. 19, p. 153–178, 2005.

KENDALL, Maurice; STUART, Alan. The advanced theory of statistics. Vol. 2. Hafner Publishing Company, New York, 1961.

KOLMOGOROV, Andrei. **Foundations of the Theory of Probability**. New York: Chelsea Publishing Company, 1956.

KOLODNY, Niko. IX—How Does Coherence Matter? *In*: WILEY ONLINE LIBRARY, 1pt3. PROCEEDINGS of the Aristotelian Society (Hardback). [*S.l.*: *s.n.*], 2007. P. 229–263.

KOLODNY, Niko. Why be rational? **Mind**, Oxford University Press, v. 114, n. 455, p. 509–563, 2005.

KRAUS, Sarit; LEHMANN, Daniel; MAGIDOR, Menachem. Nonmonotonic reasoning, preferential models and cumulative logics. **Artificial intelligence**, Elsevier, v. 44, n. 1-2, p. 167–207, 1990.

KYBURG, Henry Ely. **Probability and the logic of rational belief**. Middletown: Wesleyan University Press, 1961.

LEHMANN, Daniel; MAGIDOR, Menachem. What does a conditional knowledge base entail? **Artificial intelligence**, Elsevier, v. 55, n. 1, p. 1–60, 1992.

LEITGEB, Hannes. I—The humean thesis on belief. v. 89, n. 1, p. 143–185, 2015.

LEITGEB, Hannes. Reducing belief simpliciter to degrees of belief. **Annals of Pure and Applied Logic**, Elsevier, v. 164, n. 12, p. 1338–1389, 2013.

LEITGEB, Hannes. The review paradox: On the diachronic costs of not closing rational belief under conjunction. **Noûs**, Wiley Online Library, v. 48, n. 4, p. 781–793, 2014a.

LEITGEB, Hannes. **The Stability of Belief: How Rational Belief Coheres with Probability**. New York: Oxford University Press, 2017.

LEITGEB, Hannes. The Stability Theory of Belief. **Philosophical Review**, v. 123, n. 2, p. 131–71, 2014b.

LEVI, Isaac. Acceptance revisited. *In*: BOGDAN, Radu J. (Ed.). **Local induction**. Boston: Springer, 1976. P. 1–71.

LEVI, Isaac. **For the sake of the argument: Ramsey test conditionals, inductive inference and nonmonotonic reasoning**. [*S.l.*]: Cambridge University Press, 2007.

LEVI, Isaac. Gambling with truth: An essay on induction and the aims of science, 1967a.

LEVI, Isaac. Information and inference. **Synthese**, JSTOR, p. 369–391, 1967b.

LEVI, Isaac. **The enterprise of knowledge: An essay on knowledge, credal probability, and chance**. [*S.l.*]: MIT press, 1980.

LEWIS, David. **Counterfactuals**. Malden-MA: Harvard University Press, 1973.

LEWIS, David. Why Conditionalize? *In*: LEWIS, David (Ed.). **Papers in Metaphysics and Epistemology: Volume 2**. New York: Cambridge University Press, 1999. P. 403–407.

LIN, Hanti; KELLY, Kevin T. A geo-logical solution to the lottery paradox, with applications to conditional logic. **Synthese**, Springer, v. 186, n. 2, p. 531–575, 2012a.

LIN, Hanti; KELLY, Kevin T. Beliefs, probabilities, and their coherent correspondence. *In*: DOUVEN, Igor (Ed.). **Lotteries, Knowledge, and Rational Belief: Essays On the Lottery Paradox**. Cambridge: Cambridge University Press, 2021. P. 185–222.

LIN, Hanti; KELLY, Kevin T. Propositional Reasoning that Tracks Probabilistic Reasoning. **J Philos Logic**, v. 41, n. 0, p. 957–981, 2012b.

MAHER, Patrick. **Betting on theories**. New York: Cambridge University Press, 1993.

MAKINSON, David. Bridges between classical and nonmonotonic logic. **Logic Journal of IGPL**, Oxford University Press, v. 11, n. 1, p. 69–96, 2003.

MAKINSON, David; GÄRDENFORS, Peter. Relations between the logic of theory change and nonmonotonic logic. *In*: HILPINEN, Risto; FUHRMANN, Andre; MORREAU, Michael (Ed.). **The logic of theory change**. Berlin: Springer-Verlag, 1991. P. 185–205.

MAKINSON, David C. The Paradox of the Preface. **Analysis**, JSTOR, v. 25, n. 6, p. 205–207, 1965.

MAKINSON, David C. **The scarcity of stable belief sets**. [*S.l.*: *s.n.*], 2015. Disponível em: `https://drive.google.com/file/d/0B-gWi8__vkMNT0ZHel9kQ3V1NFU/view?resourcekey=0-0jgn_1yJXt0mb_riw5w02Q..` accessed: 13.03.2023).

MOSS, Sarah. Time-slice epistemology and action under indeterminacy. *In*: GENDLER, Tamar Szabó; HAWTHORNE, John (Ed.). **Oxford Studies in Epistemology, Vol. 5**. Oxford: Oxford University Press, 2015.

NEYMAN, Jerzy. “Inductive Behavior” as a Basic Concept of Philosophy of Science. **Revue de l’Institut International de Statistique**, JSTOR, p. 7–22, 1957.

NOZICK, Robert. **The nature of rationality**. Princeton, New Jersey: Princeton University Press, 1993.

PAPINEAU, David. There are no norms of belief. **The aim of belief**, Oxford University Press Oxford, p. 64–79, 2013.

PETERSON, Martin. **An introduction to decision theory**. [*S.l.*]: Cambridge University Press, 2017.

PETTIGREW, Richard. Epistemic Utility Arguments for Probabilism. *In*: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Winter 2019. [*S.l.*]: Metaphysics Research Lab, Stanford University, 2019.

PETTIGREW, Richard. Jamesian epistemology formalised: An explication of 'The Will to Believe'. **Episteme**, Cambridge University Press, v. 13, n. 3, p. 253–268, 2016.

POPPER, Karl R. **The Logic of Scientific Discovery**. [*S.l.*]: Hutchinson, 1959.

QUINE, Willard Van Orman; ULLIAN, Joseph Silbert. **The web of belief**. New York: McGraw-Hill, Inc., 1978. v. 2.

RAMSEY, Frank. General Propositions and Causality. *In*: BRAITHWAITE, Richard B (Ed.). **The Foundations of Mathematics and other Logical Essays**. [*S.l.*]: Routledge, 1931. P. 197–250.

RAMSEY, Frank P. Truth and probability. Reprinted in. **Studies in subjective probability**, New York: Wiley, p. 61–92, 1926.

RAZ, Joseph. The myth of instrumental rationality. **J. Ethics & Soc. Phil.**, HeinOnline, v. 1, p. 1, 2005.

REITER, Raymond. Nonmonotonic Reasoning. **Annual Reviews of Computer Science**, Annual Reviews Inc., v. 2, p. 147–186, 1987.

ROSENKRANTZ, Roger D. Cognitive Decision Theory. *In*: BOGDAN, Radu J. (Ed.). **Local Induction**. Boston: Springer, 1976. P. 73–92.

ROTT, Hans. **Change, choice and inference: A study of belief revision and nonmonotonic reasoning**. New York: Clarendon Press, 2001.

ROTT, Hans. Stability and scepticism in the modelling of doxastic states: Probabilities and plain beliefs. **Minds and Machines**, Springer, v. 27, n. 1, p. 167–197, 2017.

RUI, Matheus. Compatibilizando Teoria da Ação e Racionalidade Prática a Partir do Conceito de “Intenção”. **Analytica-Revista de Filosofia**, v. 22, n. 1, p. 177–200, 2018.

RUI, Matheus. What is the aim of models in formal epistemology? **Principia: an international journal of epistemology**, v. 26, n. 1, p. 135–152, 2022.

SAVAGE, Leonard J. **The foundations of statistics**. [*S.l.*]: Courier Corporation, 1972.

SCANLON, Thomas M. Structural irrationality. **Common minds: Themes from the philosophy of Philip Pettit**, Oxford: Oxford University Press, p. 84–103, 2007.

SCHURZ, Gerhard. Impossibility results for rational belief. **Noûs**, Wiley Online Library, v. 53, n. 1, p. 134–159, 2019.

SEARLE, John. **Intentionality: An essay in the philosophy of mind**. [*S.l.*]: Cambridge university press, 1983.

SHAFER, Glenn. **A mathematical theory of evidence**. [*S.l.*]: Princeton university press, 1976. v. 42.

SHEAR, Ted; FITELSON, Branden. Two approaches to belief revision. **Erkenntnis**, Springer, v. 84, n. 3, p. 487–518, 2019.

SHOHAM, Yoav. **Reasoning about change: time and causation from the standpoint of artificial intelligence**. Cambridge-MA: Yale University, 1987.

SKORUPSKI, John. What is normativity? **Disputatio**, Sciendo, v. 2, n. 23, p. 247–269, 2007.

SKYRMS, Brian. Causal necessity: a pragmatic investigation of the necessity of laws. Yale University Press, London, 1980.

SMITH, Michael. Meta-ethics. **The Oxford handbook of contemporary philosophy**, Oxford University Press Oxford, p. 3–30, 2005.

SPOHN, Wolfgang. A survey of ranking theory. *In*: HUBER, Franz; SCHMIDT-PETRI, Christoph *et al.* (Ed.). **Degrees of belief**. [*S.l.*]: Springer, 2009. P. 185–228.

STALNAKER, Robert. A note on non-monotonic modal logic. **Artificial Intelligence**, Elsevier, v. 64, n. 2, p. 183–196, 1993.

STALNAKER, Robert. **Inquiry**. Cambridge-MA: The MIT Press, 1984.

STALNAKER, Robert. What is a nonmonotonic consequence relation? **Fundamenta Informaticae**, IOS Press, v. 21, n. 1-2, p. 7–21, 1994.

STEELE, Katie; STEFÁNSSON, H. Orri. Decision Theory. *In*: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Winter 2020. [*S.l.*]: Metaphysics Research Lab, Stanford University, 2020.

TITELBAUM, Michael. ForthcomingNormative Modeling. *In*: HORVATH, J. (Ed.). **Methods in Analytic Philosophy: A Contemporary Reader**. [*S.l.*]: The PhilPapers Foundation, no prelo.

TITELBAUM, Michael. **Fundamentals of Bayesian Epistemology 1: Introducing Credences**. New York: Oxford University Press, 2022a.

TITELBAUM, Michael. **Fundamentals of Bayesian Epistemology 2: Arguments, Challenges, Alternatives**. New York: Oxford University Press, 2022b.

TITELBAUM, Michael. **Quitting certainties: A Bayesian framework modeling degrees of belief**. [*S.l.*]: Oxford University Press, 2013.

TVERSKY, Amos; KAHNEMAN, Daniel. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. **Psychological review**, American Psychological Association, v. 90, n. 4, p. 293, 1983.

VAN FRAASSEN, Bas C. Belief and the Will. **The Journal of Philosophy**, v. 81, n. 5, p. 235–256, 1984.

VAN FRAASSEN, Bas C. Fine-grained opinion, probability, and the logic of full belief. **Journal of Philosophical logic**, Springer, v. 24, n. 4, p. 349–377, 1995.

VINEBERG, Susan. Dutch Book Arguments. *In*: ZALTA, Edward N.; NODELMAN, Uri (Ed.). **The Stanford Encyclopedia of Philosophy**. Fall 2022. [*S.l.*]: Metaphysics Research Lab, Stanford University, 2022.

WALTMAN, Paul. **Deterministic threshold models in the theory of epidemics**. [*S.l.*]: Springer Science & Business Media, 2013. v. 1.

WEDGWOOD, Ralph. Internalism explained. **Philosophy and phenomenological research**, v. 65, n. 2, p. 349–369, 2002a.

WEDGWOOD, Ralph. The aim of belief. **Philosophical perspectives**, JSTOR, v. 16, p. 267–297, 2002b.

WEDGWOOD, Ralph. **The value of rationality**. New York: Oxford University Press, 2017.

WEISSTEIN, Eric. **Distance From MathWorld–A Wolfram Web Resource**. Disponível em: `https://mathworld.wolfram.com/Distance.html`. accessed: 26.02.2021).

WHEELER, Gregory. Formal epistemology. *In*: CULLISON, Andrew (Ed.). **The Continuum companion to epistemology**. London: Bloomsbury Publishing, 2012. P. 227–247.

WILLIAMSON, Timothy. Model-building in philosophy. *In*: BLACKFORD, R; BRODERICK, D (Ed.). **Philosophy's future: The problem of philosophical Progress**. Malden-MA: Wiley Blackwell, 2017. P. 159–172.

WRIGHT, Georg Henrik von. Norm and action: a logical enquiry. Routledge & Kegan Paul, 1963.

YAP, Audrey. Idealization, epistemic logic, and epistemology. **Synthese**, Springer, v. 191, n. 14, p. 3351–3366, 2014.

Apêndices

APÊNDICE A – POR QUE PRECISAMOS DE MODELOS?

Neste apêndice pretendo apresentar o modo como podemos entender a formalização de teorias em epistemologia, focando em (i) o papel da idealização nos modelos científicos; (ii) o conflito entre agentes ideais e reais na epistemologia; (iii) o papel dos modelos em teorias sobre a racionalidade; e (iv) a falta de um limiar claro entre o que é o fenômeno a ser modelado e o que é o próprio modelo em epistemologia[1].

CONSIDERAÇÕES INICIAIS

É bem aceito atualmente que modelos formais desempenham um papel relevante no trabalho científico. Isso tem sido o caso já por alguns séculos, principalmente nas ciências naturais como a Física e a Química, e mais recentemente para ciências sociais como a economia e política. Na filosofia, somente nas últimas décadas a discussão sobre o uso de modelos formais como uma ferramenta filosófica genuína tem recebido a devida atenção (embora o uso de ferramentas formais na atividade filosófica é uma prática de longa data). Aqui estarei preocupado com o uso de modelos formais em um campo filosófico específico, nomeadamente, a epistemologia da racionalidade.

Aqui será discutido a razão pela qual, apesar das controvérsias, o uso de modelos formais no empreendimento epistemológico pode impulsionar nossa pequena caixa de ferramentas metodológicas ao fazer epistemologia. Nas seções 2, 3 e 4, apresento brevemente as principais linhas de críticas e problemas para um modelador formal. Na última seção, pretendo oferecer alguns *insights* que podem beneficiar o presente tema. Entre as contribuições teóricas, sustento as seguintes teses: (i) modelos formais são ferramentas úteis para prever consequências de suposições normativas sobre o que é intuitivamente exigido da racionalidade, e (ii) na medida em que uma teoria da racionalidade é normativa em virtude de ser instrumentalista e visar a verdade, modelos formais são ferramentas meios-fins, de modo que, para a racionalidade, modelos matemáticos são recursos para maximizar a verdade em estados doxásticos.

MODELOS FORMAIS NAS CIÊNCIAS

O entendimento científico do que significa ser um “modelo” é extenso. Inclui algoritmos computacionais, estruturas matemáticas, figuras, e até mesmo objetos concretos. Aqui me restrinjo apenas aos modelos *formais*. Por “formal” quero dizer que as propriedades do modelo são matematicamente estipuladas. Em outras palavras, tais modelos são construídos sobre estruturas matemáticas, as quais podem incluir

[1] O conteúdo apresentado aqui consiste basicamente em um compilado (e uma tradução para o português) das contribuições que foram publicadas em Rui (2022), e apresentadas no *12th Principia International Symposium: Models and Modeling in the Sciences*, no ano de 2021. Agradeço o Prof. Dr. Ivan Ferreira da Cunha pela organização desse dossiê publicado na revista *Principia*.

Lógica, números, conjuntos, funções, vetores, etc. Tradicionalmente, parte do trabalho do cientista consiste em descrever o modelo e conectá-lo a um fenômeno no mundo por meio de princípios teóricos. O uso de ferramentaria matemática permite ao pesquisador considerar cenários mais complexos, que envolvem diversas variáveis. Alguns modelos são desenvolvidos para fazer previsões, outros para descrever um fenômeno, ou simplesmente para aperfeiçoar a explicação de algum evento.

Resumidamente, o trabalho de modelagem possui duas partes fundamentais: a estrutura de modelagem e a interpretação da estrutura. Saber com interpretar deve ser uma competência essencial do modelador. Primeiramente, precisamos interpretar como os dados entram no modelo, e em seguida saber como interpretar as conclusões do modelo. Estas são tarefas distintas. Uma interpretação clara permite a distinção precisa entre o que é elemento do modelo e o que é elemento do mundo. Ainda, modelos são estruturas abstratas, eles não exigem que todos seus elementos estejam no mundo, a menos que sua interpretação assim o exija.

Agora sendo mais objetivo, vamos analisar um par de exemplos conhecidos da prática científica. Um dos modelos epidemiológicos mais rotineiramente utilizado para compreender a propagação de uma infecção dentro de uma população é o SEIR (Suscetível-Exposto-Infectado-Recuperado). Particionando a população em quatro grupos exclusivos, tal modelo é construído a partir do seguinte sistema de equações diferenciais:

$$\frac{dS}{dt} = \mu N - \beta(I)S - \mu S \tag{14}$$
$$\frac{dE}{dt} = \beta(I)S - (\mu + \epsilon)E \tag{15}$$
$$\frac{dI}{dt} = \epsilon E - (\gamma + \mu)I \tag{16}$$
$$\frac{dR}{dt} = \gamma I - \mu R \tag{17}$$

Este conjunto de equações diferenciais é utilizado para relacionar o modelo com as taxas $\beta, \gamma, \mu, \epsilon$ onde a população migra de um grupo para o outro. Para ser mais preciso, cada equação diferencial do lado esquerdo pode ser colocada em um gráfico (e aqui podemos começar a entender a famosa fala sobre "achar a curva"), e desse modo o modelo pode prever o quão grave pode ser uma epidemia. Cada equação, e consequentemente, cada gráfico, é determinado pelos parâmetros denotadas pelas letras gregas. β é a taxa de infecção (para quantas pessoas um infectado pode transmitir a doença), γ é o período de infectividade, μ é a taxa de mortalidade, e ϵ é o período de incubação (o tempo entre a infecção e o aparecimento dos sintomas). O número total de indivíduos na população é denotado pela letra N, de tal modo que $S + E + I + R = N$.

Como é de se esperar, não aprofundarei os detalhes teóricos aqui. O objetivo do exemplo é chamar a atenção para dois aspectos do modelo SEIR. Primeiramente, o modelo estipula uma relação entre premissas de tal modo que, se todos os dados de entrada são acurados, as conclusões ocorrerão. Isso pode ser resumido no famoso mandamento estatístico *"Garbage In, Garbage Out!"* (ou seja, se entra lixo, sai lixo). Assim, o modelo aqui não é apenas uma ferramenta para fazer previsões, mas também serve como para avaliar a acurácia dos dados iniciais inseridos no modelo na medida em que as previsões venham a se confirmar. O segundo aspecto que quero enfatizar aqui é a idealização, ou natureza irrealista, de um modelo como esse. Durante a construção do modelo, ao menos três suposições falsas são feitas. A primeira consiste no fato de que o número de indivíduos em cada grupo é tratado como contínuo ao invés de discreto. A segunda falsidade consiste na suposição de que o número da população é constante, ignorando nascimentos e mortes. E terceiro, é suposto que a taxa de recuperação é igualmente provável entre os infectados (WALTMAN, 2013).

O próximo, e último, exemplo de modelo formal no campo científico é um mais controverso: o modelo de preços na teoria microeconômica. Na ciência econômica, mais precisamente na microeconomia, "preço" é basicamente definido como uma função de oferta e demanda, determinada pela relação entre a quantidade e preço de um determinado bem. Tudo isso é feito por meio de modelagem formal. Brevemente, o objetivo da microeconomia é estudar o comportamento de agentes econômicos nas tomadas de decisões individuais e coletivas. Uma suposição de fundo nesse modelo é a controversa suposição do *homo economicus*. Na teoria microeconômica, esse suposto agente é um agente da teoria da decisão, e isso significa que é um agente que maximiza a utilidade esperada de suas ações de acordo com suas preferências racionais (ou seja, preferências completas e transitivas).

No auge do debate, o economista Milton Friedman publicou um livro em defesa das suposições irrealistas em modelos econômicos. Para ele, um modelo deve ser julgado pela sua capacidade de realizar previsões precisas, e não pela acurácia de suas premissas. Nas palavras do autor,

> Hipóteses realmente importantes e significativas podem ser descobertas como contendo "suposições" que são representações descritivas altamente imprecisas da realidade e, em geral, quanto mais significativa a teoria, mais suposições irrealistas (nesse sentido). A razão é simples. Uma hipótese é importante se ela "explica" muito com pouco, ou seja, se ela abstrai os elementos comuns e cruciais da massa de circunstâncias complexas e detalhadas ao redor do fenômeno a ser explicado, e permite previsões válidas apenas a partir disso (FRIEDMAN, 1953).

O que Friedman chama aqui de "Teoria" é precisamente o que estamos chamando de modelo. Podemos observar claramente na citação o aspecto irrealista das suposições no modelo, e isso ocorre pelo bem de previsões precisas. Em resumo: um modelo formal é uma ferramenta para o objetivo final do cientista (ao menos no sentido defendido por Friedman).

Novamente, o objetivo aqui não é oferecer uma explicação detalhada sobre determinado modelo matemático em uma área específica. Acredito que o leitor já esteja relativamente familiarizado com vários exemplos semelhantes aos que foram abordados acima. Tudo que preciso é de um paralelo nas ciências. Como veremos abaixo, o uso de modelos na epistemologia algumas vezes se aproxima e outra se distancia dos casos acima, e a justificação de seu uso precisa ser feita, seja quando a metodologia filosófica se encaixa ou diverge dos casos científicos. Comecemos, então, a falar de filosofia daqui em diante.

O QUE TUDO ISSO TEM A VER COM FILOSOFIA?

Nas ciências, alguns modelos matemáticos são desenvolvidos para fazer previsões, outros para descrever fenômenos, ao simplesmente para aumentar o poder explicativo de uma teoria. Mas e na filosofia? Apesar das controvérsias, o uso de modelos formais em problemas filosóficos tem crescido nas últimas décadas. Todavia, antes de dar prosseguimento ao nosso tópico central, é importante lançar luz sobre uma prática filosófica que não é propriamente a de modelagem, embora muitas vezes seja tradada como uma.

Uma prática bastante difundida dentro da literatura filosófica contemporânea, especialmente aquela de língua inglesa, consiste em passar da descrição em palavras de um problema filosófico para seu desenvolvimento em termos de símbolos e equações. Algumas vezes, o uso de uma linguagem formal permite ao filósofo expressar relações complexas e seu uso torna seu trabalho mais preciso e compacto. Chamemos esse de o uso da linguagem formal como "abreviação da linguagem natural". Durante o século XX, o uso da linguagem formal como abreviação tornou-se padrão na epistemologia. Em um artigo memorável, "*Is Justified True Belief Knowledge?*" Edmund Gettier retratou as condições de satisfação para o conhecimento como "*S* sabe que *p*, *sse*, *p* é verdadeira, [...]". Foi provavelmente devido ao uso da linguagem formal como abreviação que o autor brilhantemente mudou a história da epistemologia em apenas três páginas. Contudo, isso não é precisamente um exemplo de modelagem na filosofia.

Pode ser o caso que filósofos utilizem uma linguagem formal para descrever não um problema particular, mas sim um *modelo* formal. Nesse sentido, o uso de modelos na filosofia vai além do mero uso da linguagem formal como abreviação. Como bem afirma Timothy Williamson, "a clareza matemática na descrição torna o estudo direto do modelos mais fácil do que o estudo do fenômeno propriamente" (2017, p. 161). Nosso objetivo aqui será investigar os debates em torno do papel de modelos formais em um domínio filosófico específico, mais precisamente, sobre a epistemologia da racionalidade. Pretendo mostrar que o trabalho epistemológico pode ser otimizado ao se adicionar a atividade de modelagem como uma ferramenta.

Os estudos de caso aqui serão modelos típicos da epistemologia formal, e mais especificamente sobre teorias formais da racionalidade. Primeiramente, racionalidade epistêmica é sobre a racionalidade da crença (*simpliciter*) e da probabilidade subjetiva, como podemos ver ao longo do presente trabalho.

Para o presente propósito, tomarei os postulados lógicos-matemáticos qualitativos e quantitativos para a racionalidade doxástica, conforme apresentados no Capítulo 1, como descrevendo um modelo formal de racionalidade para tais estados. Ao invés de me aprofundar em tais modelos e seus usos na epistemologia (algo que faço ao longo da Tese), este apêndice se dedicará exclusivamente no aspecto metodológico muitas vezes implícitos no uso desses postulados. Não é difícil notar que os modelos formais para crença e probabilidade subjetiva não seguem o mesmo padrão dos modelos científicos vistos acima. Mais do que isso, modelos epistemológicos não são ferramentas para se confirmar hipóteses empíricas, e nem mesmo para confirmar suas premissas por meios de previsões posteriormente confirmadas. No geral, tais modelos também não servem para prever como seres humanos agirão e, em alguns casos, eles inclusive contradizem alguns padrões comportamentais humanos. Ainda, todo os modelos de racionalidade epistêmica trabalhados ao longo da Tese tomam agentes epistêmicos como idealizados, algo como um super-herói epistêmico. Então devemos nos perguntar: será que somos capazes de explicar por que modelos formais são relevantes no trabalho epistemológico?

Além do mais, é bem conhecido o fato de que tais modelos não representam o modo como seres humanos se comportam na realidade. Mas alguém pode alegar que esses tipos de modelos epistêmicos são estritamente normativos, e isso significa que eles fornecem um conjunto de regras que regras que agentes humanos “deveriam” seguir. Todavia, estes são requisitos bastante duros para agentes epistêmicos. Isso fica claro na medida em que os modelos epistêmicos apresentado no primeiro capítulo da Tese exigem dos agentes que eles sejam logicamente oniscientes, isto é, que eles acreditem com confiança máxima em toda verdade lógica e acredite em toda proposição acarretada por seu conjunto atual de crenças. Mas antes de começar a analisar o estatuto normativo de modelos formais na epistemologia, gostaria de tornar mais precisa a compreensão do conceito de “idealização” que será bastante utilizado aqui.

O QUE SÃO MODELOS IDEALIZADOS?

Se você já leu um artigo sobre Epistemologia Formal (ou se você já leu toda a presente Tese antes de chegar neste Apêndice), você certamente já se deparou com a noção de “agentes idealizados”. Essa é uma suposição comum na grande maioria dos trabalhos formais em epistemologia. Para alguns epistemólogos, esse tipo de suposição soa como uma "bizarrice", enquanto para outros isso é tão óbvio que não merece nem mesmo ser explicitamente mencionado no texto. Mas o que é realmente assumido quando um epistemólogo considera um agente idealizado em sua teoria? A seguir, direcionarei algumas possível respostas para essa pergunta. De acordo com Michael Titelbaum no artigo *"Normative Modeling"* (no prelo), um tipo de idealização simplifica para o bem do cálculo matemático - o que ocorre nos exemplos dos modelos científicos vistos acima. Um agente ideal, por outro lado, é ideal no sentido de fazer o que agentes não-ideais deveriam fazer, ou é melhor que um agente não-ideal em uma dimensão normativa. Tal distinção relevante também aparece em David Christensen (2004):

> [...] tal modelo pode idealizar normativamente: ele pode visar representar idealmente crenças racionais - crenças que exibem um tipo de perfeição da qual humanos não são capazes. Mas ele também pode idealizar de um modo do qual inúmeros modelos descritivos idealizam: ele pode atribuir um número a uma quantidade cuja aplicação em casos reais não é totalmente precisa. (CHRISTENSEN, 2004, p. 145)

Feita as devidas considerações, examinarei agora tal distinção relevante levando em consideração os aspectos metodológicos geralmente assumidos na prática filosófica e o contraste com os casos mais comuns nas ciências empíricas.

Idealização como Simplificação

Idealização está muitas vezes conectada com simplicidade, e essa simplicidade pode ser compreendida de duas formas: como uma simplificação matemática, ou como uma simplificação do fenômeno. Muitas vezes esses dois alvos aparecem simultaneamente. Começando pelo primeiro caso, a idealização por conveniência matemática é mais comum do que o leitor possa imaginar. Retomando nossos exemplos iniciais, grande parte dos modelos científicos simplificam por conveniência matemática. E isso também ocorre com nossos modelos epistêmicos. Agora um paralelo pode ser traçado: na medida em que o modelo epidemiológico SEIR assume que o número de pessoas pode ser representado por qualquer número real (e sabemos que o número de pessoas em uma população é sempre discreto), parece-nos uma suposição plenamente viável aceitar que bayesianos representem probabilidades subjetivas por uma linha contínua de números reais. Mas tal comparação imediatamente levanta o questionamento: seria este um paralelo válido?

O requisito de que graus de crença sejam representados por funções real-valoradas (funções de probabilidade) parece-nos um tipo de idealização por conveniência matemática: não há nada de mais ou menos racional em representar agentes epistêmicos com funções contínuas ao invés de discretas, ou com estados doxástico cardinais ao invés de ordinais. Nesse sentido, alguns aspectos da estrutura do modelo são construídos apenas para o bem da adequação matemática. Um caso interessante aqui é a existência de teorias alternativas para a racionalidade dos estados doxásticos quantitativos, diferentes do modelo bayesiano de probabilidades subjetivas. Alguns filósofos[2], com elegantes e sofisticadas construções lógico-matemática, têm procurado por maneiras alternativas de representar o mesmo fenômeno doxástico de crenças graduais. Em alguns casos, pequenas diferenças na compreensão do fenômeno relevante pode conduzir o teórico a estruturas completamente distintas. Deixando de lado o debate substancial sobre as suposições normativas de cada uma dessas teorias (o que não é meu foco aqui), as distintas estruturas matemática por trás de cada teoria possui um propósito instrumental, cuja intenção é modelar uma concepção mais apropriada de crença gradual[3].

Caminhando agora em direção ao ao segundo tipo de idealização como simplificação, temos o caso de simplificação do fenômeno. Em alguns casos, simplificamos algo por ser “suficientemente próximo”. Isso ocorre geralmente quando o tópico é muito complexo para tratá-lo diretamente. O exemplo do *homo economicus* na teoria Microeconômica parece se encaixar muito bem aqui, e acredito eu, se encaixa também em muitos dos casos de teorias filosóficas da racionalidade. De acordo com Sven Hansson (2018, p. 16), “[a] razão do porquê simplificamos-idealizamos é que a questão filosófica é comumente tão complexa que uma tentativa de cobrir todos os aspectos sobrecarregaria o modelo ao ponto de torná-lo inútil”.

Em tal caso, a teoria que se propõe a cobrir todos os aspectos do fenômeno relevante seria inatingível em nosso contexto teórico atual. Por essa razão, o modelador desvincula de sua teoria variáveis importantes e investiga pontos isolados. Uma possível consequência disso é que o resultado pode ser um modelo que se pareça bastante diferente dos fatos sob investigação. Seguindo Hanson, tal desvio deve sempre ser julgado relativamente ao propósito do modelo, e como ele é utilizado: se qualquer característica pertinente for “perdida” durante a idealização, devemos considerar o quanto perderíamos em simplicidade se tal característica for incluída. Nas palavras do autor, “modelagem filosófica ou científica é sempre uma troca entre simplicidade e fidelidade ao original” (HANSSON, 2018, p. 164).

Algumas vezes o emprego de simplificação em problemas filosóficos é demasiadamente ingrato. Um caso bastante interessante é de um dos mais tradicionais modelo formais na epistemologia, o da Lógica Epistêmica. De um modo bastante perspicaz, Audrey Yap afirmar que “lógica epistêmica ou falha em ser lógica ou falha em ser epistêmica” (YAP, 2014, p. 03). Para alguns, a busca por uma lógica epistêmica é algo inatingível por sua própria natureza. Para outros, tais desvios “não verdadeiramente epistêmicos” na Lógica Epistêmica são apenas efeitos colaterais do empreendimento

[2] Ver, por exemplo Spohn (2009) e Shafer (1976).
[3] Um amplo debate sobre tais concepções alternativas na representação formal de crença gradual pode ser encontrado em Huber (2009) e Genin e Huber (2020).

de ser uma lógica. Ainda sobre isso, para finalizar, acredito que muitas das rusgas filosófica-lógica-epistêmicas podem ser solucionadas a partir de uma melhor compreensão sobre o papel dos modelos formais na distinção sobre o que está no mundo (fenômeno) e o que está no modelo, e sobre os tipos de idealização subjacentes em tais teorias.

Idealização como perfeição cognitiva

Como discutido previamente, parte do formalismo bayesiano pode ser tratado como um tipo de idealização como simplificação. No entanto, a exigência de que os agentes devem ter crenças graduais em acordo com os axiomas probabilísticos parece carregar um componente normativo: agentes que descumprem tal requisito em seus processo de raciocínio podem sofrer algum tipo de perda. Tal "perda" tem sido muito investigada pela literatura bayesiana (como pode ser visto em vários momentos ao longo da presente Tese), seja uma perda monetária (a partir de um argumento *Dutch Book*), ou uma perda epistêmica (a partir de um argumento de dominância por acurácia)[4]. Desse modo, um agente epistêmico que satisfaz os axiomas da teoria da probabilidade é ideal no sentido de fazer o que agentes não-ideais "deveriam" fazer. E o tipo relevante de idealização aqui é um caso de perfeição cognitiva.

Idealização como perfeição cognitiva é um tipo de idealização que dificilmente encontramos nos modelos das ciências empíricas[5]. Lidar com conceitos normativos é reconhecidamente uma tarefa da filosofia. Nesse sentido, é comum o uso de conceitos normativos no trabalho epistêmico, e é aqui onde nosso trabalho se distância do trabalho científico. Sobre a distinção entre uma investigação normativa e científica, Gilbert Harman afirma:

> Realmente, teorias normativas e descritivas do raciocínio estão intimamente relacionadas. [...] é difícil surgir com princípios normativos convincentes sem levar em consideração como as pessoas de fato raciocinam, o que é competência de uma teoria descritiva. Por outro lado, parece que qualquer teoria descritiva deve envolver uma certa quantidade de idealização, e idealização é sempre normativa em alguma medida. (HARMAN, 1986, p. 07)

4 Para o leitor não familiarizado com tal literatura, recomendo Vineberg (2022) e Pettigrew (2019).

5 O leitor agora pode estar pensando sobre a suposição do *homo economicus*, tal como vimso anteriormente. Mas mesmo naquele cenário, se assumirmos a noção de modelagem microeconômica de Milton Friedman, a suposição do *homo economicus* possui um papel puramente instrumental, como um meio de facilitar previsões. Nesse sentido, acredito que a fronteira entre empreendimentos científicos e filosóficos pode ser traçada pelo modo como cada campo do saber lida com o tipo de suposição normativa em seu modelo: enquanto ciências empíricas utilizam de suposições normativas instrumentalmente, é na filosofia que tal suposição assume o papel de alvo primário de investigação. Mas reconheço aqui a existência de tópicos que não somos capazes de encaixar perfeitamente nessa distinção, como, por exemplo, algumas interpretações sobre Teoria dos Jogos. Um debate formidável sobre Teoria dos Jogos e concepções "Normativas *vs* Descritivas" pode ser encontrada em Binmore (1997).

Note que tal tipo de idealização não é exclusividade da epistemologia formal, mas também ocorre nas discussões em epistemologia tradicional e em outros campos filosóficos. Mesmo em um "clube" de epistemólogos tradicionais, geralmente assume-se uma intrínseca conexão entre a suposição de um agente ideal e a normatividade da epistemologia.

O conflito sobre como devemos fazer teoria sobre o conhecimento, seja normativa ou descritiva, está no pano de fundo de diversas discussões sobre epistemologia da racionalidade: sempre há uma troca teórica entre considerar como as pessoas de fato raciocinam e considerar como um agente ideal deveria raciocinar baseado em "princípios normativos". Sobre esse tipo de troca, Harman afirma:

> Como devemos começar a desvendar o que são tais princípios de revisão? Parece haver duas abordagens possíveis. Podemos começar considerando como as pessoas *de fato* raciocinam, tentando desvendar quais princípios elas *realmente seguem*. Ou podemos começar por nossas "intuições" como *críticos* do raciocínio. (HARMAN, 1986, p. 09)

Nesse sentido, "princípios normativos" seriam uma espécie de "intuição filosófica" (embora eu não seja capaz aqui de especificar exatamente o que seria isso). Seguindo o texto de Harman citado acima, chamarei tal conflito (sobre como devemos começar teorizando) de *O Dilema de Harman*. Isso será central em nosso próximo tópico, e retornarei a tal distinção em breve.

De volta a noção de idealização, muitas vezes é difícil decidir se uma idealização é feita visando conveniência matemática, simplificação do fenômeno relevante, ou para mostrar o que um agente ideal deveria fazer. Um caso paradigmático é o problema da "Onisciência Lógica" na literatura padrão da Lógica Epistêmica. O fato de seus agentes epistêmicos claramente possuírem capacidades que não não carregadas por seres humanos levou alguns teóricos a argumentar contra a Lógica Epistêmica como uma teoria sobre conhecedores reais. Uma possível resposta a tal crítica seria argumentar que onisciência lógica é meramente um efeito colateral da modelagem formal (e sua idealização) na Lógica Epistêmica. Mas não está claro qual tipo de idealização subjaz o requisito de fechamento dedutivo (o principal responsável pelo problema da onisciência lógica), se é de um tipo normativo com agentes cognitivos de capacidade ilimitada, ou uma simplificação para o bem do empreendimento teórico.

No restante deste apêndice que o uso de uma abordagem metodológica que emprega modelos formais em problemas epistêmicos é um ponto sem volta. Todavia, fazer modelagem epistêmica de modo apropriado requer uma compreensão mais precisa de seu papel como um modelo formal. Para realizar isso, precisamos considerar a distinção acima sobre tipos diferentes de idealização: diferentes motivações devem ser levadas em consideração quando queremos criticar um modelo. Nesse sentido, um modelo não deve ser criticado como inadequado simplesmente por se utilizar de idealização ou suposições falsas. E finalmente, pretendo chamar a atenção para a suposição normativa sobre uma teoria específica da racionalidade e as características de agentes ideais em seus modelos.

POR QUE FAZER MODELAGEM EPISTÊMICA?

Nesta seção, ofereço duas razões que considero centrais em favor da realização de modelagem formal na epistemologia. A primeira é motivada por um elemento central que distingue modelagem epistêmica e científica: a suposição normativa. Meu propósito será reforçar a tese de que alguns problemas epistêmicos envolvendo suposições normativas, especialmente aqueles envolvendo epistemologia da racionalidade, não podem ser adequadamente compreendidos sem algum tipo de modelagem formal. E, para finalizar, a segunda razão para se fazer modelagem epistêmica é motivada por uma visão específica da racionalidade: a teoria instrumental. Como veremos, serão apresentadas algumas razões para se tomar a racionalidade como uma ferramenta de cálculos de meios-fins, e pretendo explicar como a atividade de modelagem epistêmica pode nos auxiliar na busca por uma teoria da racionalidade. De modo geral, acredito que as considerações feitas no restante deste apêndice representam um progresso na (escassa) literatura recente sobre o tema.

Modelos Formais para Teorias Normativas

Frente às objeções de que agentes reais falham em cumprir com requisitos de grande parte dos modelos epistêmicos existentes, alguns epistemólogos formais têm argumentado que modelos idealizados são ferramentas para teorias normativas, e são muito mais sobre como agente “deveriam” agir do que sobre como eles de fato agem. De acordo com Titelbaum (no prelo) e Colyvan (2013), teorias epistêmicas da crença racional são normativas no sentido de agentes ideais são “epistemicamente melhores” do que agentes reais. Assim, modelos bayesianos e lógicos (como os apresentados ao longo da Tese) são modelos sobre “fatos normativos”. De acordo com Colyvan (2013, p. 1340), como parte de teorias normativas, modelos formais devem ser tomados como uma prescrição de como devemos raciocinar, organizar nossas crenças, entre outras coisas epistemicamente relevantes.

Um possível modo de compreender o que está sendo aqui chamado de “fatos normativos” é tomando-os como, citando Harman (1986, p. 09) novamente, “intuições” oriundas de nossa capacidade como “críticos do raciocínio”. Nesta seção, de uma maneira próxima a Titelbaum (no prelo) e Colyvan (2013), aceitarei que modelos epistêmicos formais são modelos para algum tipo de teoria normativa sobre a racionalidade. E assumindo que teorias da racionalidade são normativas, em alguma medida, seus modelos devem incluir em suas entradas algum tipo de “intuição normativa” (advinda de críticos do raciocínio). E aqui podemos notar que a modelagem epistêmica é bastante similar a modelagem científica; ela se parece a científica, mas feita com fatos normativos. No entanto, diferentemente do modo científico de retirar conclusões dos dados introduzidos, a modelagem epistêmica é também capaz de avaliar a viabilidade dos fatos de entrada no modelo, e isso pode ser feito ao se prestar atenção as consequências que são resultado do que foi introduzido no modelo. Como Wheeler (2012, p. 233) afirma, "a descoberta de características robustas de um problema pode remodelar nossas intuições. E precisamente desse modo a epistemologia formal pode ser utilizada para treinar intuições filosóficas". Resumindo, o processo de modelagem formal pode calibrar nossas intuições sobre a racionalidade epistêmica.

Para reforçar o ponto do parágrafo anterior, trago aqui a caso do paradoxo da loteria. No primeiro capítulo da Tese, Seção 1.3.3, apresentamos formalmente o paradoxo da loteria, resultado da combinação da teoria bayesiana para probabilidade subjetiva, do fechamento dedutivo para crença, e da Tese Lockeana como princípio ponte. No entanto, o referido paradoxo pode ser apresentado de uma maneira totalmente informal, sem a utilização de números, fórmulas, ou regras lógicas (como a lei De Morgan). Para compreender o caráter paradoxal, basta que pensemos em uma loteria hipotética, sabidamente justa, da qual você é portador de um único bilhete. Dado que a loteria é relativamente grande, você possui uma baixa confiança de que seu bilhete é o vencedor, o que torna razoável sua crença de que não ganhará na loteria. Mas como seu bilhete não tem nada de especial em relação aos outros, você poderia acreditar em cada um dos bilhetes particularmente como sendo bilhetes perdedores. Assim, você termina acreditando que nenhum bilhete é vencedor, mas que um deles certamente é vencedor.

Como o leitor deve ter percebido, a apresentação acima não exige nenhuma capacidade formal significativa para a compreensão do caráter paradoxal do caso. Todavia, a mera constatação da existência de um paradoxo não nos parece suficiente: queremos entender o que fundamenta tal paradoxo e, se possível, oferecer uma solução (ninguém gosta de ficar preso em paradoxos!). O ponto central aqui é mostrar como problemas aparentemente simples possuem uma estrutura extremamente complexa de fundo, e isso foi observado durante toda a presente Tese ao explorar várias possíveis respostas ao paradoxo. O que quero dizer é que, apesar de plenamente compreensível de modo informal, é impossível entender o que realmente está acontecendo no paradoxo da loteria sem utilizar de algumas das teorias formais da racionalidade da crença. Como vimos ao longo da Tese, o que o paradoxo nos mostra é que existe um profundo problema com nossas melhores teorias sobre a racionalidade da crença. É importante enfatizar que o que está sendo pressuposto aqui é que o que está errado no caso do paradoxo da loteria é o que Hempel (1962) chamou de "não conjuntividade da sistematização estatística": a probabilidade de um conjunto de proposições conjuntivas cai na medida em que adicionamos novas proposições ao conjunto. E tal problema não pode ser simplesmente respondido informalmente (como o leitor deve ter notado ao longo das mais de 150 páginas que precedem este apêndice). Portanto, responder ao paradoxo da loteria exige um estudo profundo sobre os modelos lógicos e bayesiano sobre a racionalidade doxástica.

O problema de olhar apenas informalmente para o paradoxo da loteria é que dificilmente seríamos capazes de isolar os aspectos relevantes e compreender o que está acontecendo exatamente. O trabalho de modelagem epistêmica nos permite fatiar o problema em três partes distintas: o modelo de crença (qualitativa) racional, o modelo bayesiano de probabilidade subjetiva, e um modelo para o princípio ponte relevante. Nesse sentido, modelos epistêmicos são relativamente semelhantes aos modelos científicos, mas com fatos normativos sobre o que é exigido pela racionalidade. Em poucas palavras, modelos formais podem nos ajudar a prever consequências e lançar luz sobre teorias da racionalidade da crença. Sobre isso, Colyvan afirma:

> No caso de teorias normativas, dentre os frutos podemos incluir: mexer com nossas intuições sobre o que fazer em certas situações, lançar luz sobre situações onde a intuição falha, e talvez facilitar a construção de elegantes teoremas da representação. (COLYVAN, 2013, p. 1347)

Agora podemos nos perguntar: dada a incompatibilidade entre os três (acima citados) modelos que dão origem ao paradoxo da loteria, qual deles deve ser rejeitado? Algum deles deve ser rejeitado? A resposta a tal questionamento depende do que devemos considerar um critério de avaliação para cada modelo. Em outras palavras: o que faz de uma teoria da racionalidade uma “boa teoria”? E aqui devemos voltar ao Dilema de Harman, da seção anterior: fazer teoria da racionalidade exige saber manipular o conflito entre como as pessoas agem de fato e o que nossas intuições filosóficas nos dizem sobre o que é um requisito racional. Na medida em que uma teoria da racionalidade epistêmica, como aqui assumido, possui um componente normativo, em algum momento nossas intuições devem entrar no modelo. No caso lotérico, podemos questionar, para citar apenas dois exemplos, qual é o estatuto normativo do requisito de consistência (isto é, de que nossas crenças não devem contradizer uma às outras), ou podemos questionar o estatuto normativo do requisito de conjuntividade (isto é, de que todas nossas crenças podem ser colocadas lado a lado em uma grande conjunção de uma única proposição). Dado seu apelo intuitivo, somente um filósofo mesmo para questionar algo como tais exigências. Mas o que são tais intuições normativas e como justificá-las? Não sou capaz de responder essa dura questão aqui, mas espero que, nas páginas seguintes, eu seja capaz de direcionar alguns pontos interessantes sobre intuições normativas e a prática de modelagem epistêmica.

Em muitos casos, não somos capazes de distinguir se o requisito normativo em uma teoria da racionalidade é o efeito de um intuição filosófica ou se a respectiva intuição é um efeito do modelo. Por exemplo, podemos fazer as seguintes perguntas:

O que justifica o princípio de não-contradição na Lógica Epistêmica? É o modelo que justifica o requisito ou o requisito que justifica o modelo? É uma intuição filosófica que justifica a lógica clássica ou a lógica que justifica a intuição? É claro que essas são questões bastante complexas e fascinantes que perpassam a epistemologia, filosofia da lógica, da ciência e da matemática. E tudo isso está relacionado com a questão sobre o que torna uma teoria da racionalidade epistêmica uma boa teoria.

Para tornar as coisas mais precisas, apresentarei um exemplo interessante no modelo bayesiano de probabilidades subjetivas. Chamo de "requisito de não-conjuntividade" a regra probabilística que diz que a probabilidade de dois ou mais eventos não é nunca estritamente maior que a probabilidade de apenas um deles. O fato de que agentes reais dificilmente satisfazem tal requisito deu origem ao famoso *Experimento da Falácia da Conjunção*, ou mais conhecido também por *O Problema de Linda*. Em um consagrado artigo, os psicólogos Tversky e Kahneman (1983) testaram as intuições das pessoas sobre juízos probabilísticos. O que eles encontraram foi que os participantes do experimento desviavam do requisito de não-conjuntividade quando o cenário menos provável era mais "representativo" do que o cenário mais provável e menos "representativo". O que eles concluíram foi que os participantes tendiam a seguir um tipo de "heurística" em seus raciocínios, ao invés de alguma regra probabilística. E isso não é necessariamente algo ruim. O raciocínio guiado por um procedimento heurístico é cognitivamente mais simples do que um raciocínio guiado por alguma regra bayesiana abstrata. Mas o que isso significa para nós?

O parágrafo anterior se encaixa muito bem como um exemplo do quão significativo é o Dilema de Harman em nossa decisão metodológica ao fazer teoria da racionalidade epistêmica. Ocasionalmente, fatos sobre como as pessoas raciocinam não se encaixam com nossas teorias normativas da racionalidade. Mas novamente: o que justifica uma teoria da racionalidade? Bem, para o bayesianismo, os famosos argumentos em favor do probabilismo estão fundamentados em ambas as vantagens pragmáticas, com um argumento *Dutch Book*, e epistêmica, com um argumento de dominância por acurácia. Assim, o requisito de não-conjuntividade não é meramente uma regra fundada em nossas "intuições como críticos do raciocínio", mas ele também se baseia em nossos melhores argumentos sobre como um agente bayesiano deve raciocinar para maximizar a verdade de seus estados doxásticos (ou seu bem estar na instância prática). E isso não é apenas o caso para o requisito de não-conjuntividade, mas sim para todas regras probabilísticas. O que isso significa que é requisitos bayesianos para a racionalidade não precisam estar fundamentados essencialmente em uma intuição filosófica ideal, fruto de um mente excepcional, mas sim em um possível modo de maximizar objetivos pré-estabelecidos[6].

[6] É importante mencionar que tanto o argumento *Dutch Book*, como o argumento de dominância por acurácia, possuem pressupostos que não são aceitos por todos. O leitor pode conferir Vineberg (2022) e Pettigrew (2019) para ver mais sobre tais controversas. Eu reconheço aqui que a aceitação

Modelos Formais como ferramenta de cálculo sobre meios-fins

O último tópico deste apêndice é uma tentativa de oferecer algum tipo de fechamento para algumas questões que ficaram em aberto na seção anterior, e com isso oferecer ao leitor uma razão em favor dos benefícios de se fazer modelagem epistêmica formal. Mais do que simplesmente ser um modelo para teorias normativas da racionalidade, podemos tomar os modelos lógicos e o bayesiano como ferramentas instrumentais. A seguinte analogia pode ser útil: na medida em que a Teoria da Decisão fornece ao agente um guia de ação que maximize a utilidade esperada em contextos complexos onde o agente é incapaz de "notar" (ou calcular) a melhor opção, um modelo epistêmico pode fornecer ao agente (ou para alguém que o assiste) o melhor caminho para maximizar a verdade em circunstâncias complexas. Isso significa que modelos formais na teoria da racionalidade se encaixam muito bem com a visão instrumental da racionalidade, ou seja, uma visão que toma a racionalidade como uma ferramenta para o cálculo dos melhores meios para se atingir os fins desejados.

A ideia de avaliar requisitos racionais por sua capacidade de promover a maximização de verdade não é algo necessariamente novo. Proponentes do Argumento da Utilidade Epistêmica[7] argumentam que a racionalidade de um estado epistêmico é determinada por uma função que mede a utilidade (ou desutilidade) epistêmica de possuir determinados estados doxásticos em dadas situações. Assim, requisitos racionais na teoria da utilidade epistêmica são os clássicos requisitos da teoria da utilidade que visam maximizar o valor epistêmico. E o "bem epistêmico" mais comumente assumido é a verdade, ou também a "acurácia"[8]. Como mencionado anteriormente aqui (e também ao longo da Tese), argumentos *Dutch Book* e de Dominância por Acurácia são os argumentos tipicamente utilizados para justificar regras probabilísticas para o bayesianismo, e em ambos os argumentos se baseiam em normas da teoria da utilidade, fundamentalmente sobre a noção decisória de dominância[9].

O que a Teoria da Utilidade Epistêmica nos oferece torna a noção relevante de racionalidade epistêmica algo mais palatável. Como resultado disso, o papel de agente idealizados no processo de modelagem torna-se algo melhor direcionado. O paradoxo da loteria é um bom exemplo disso. Como vimos ao longo da Tese, nas últimas décadas o debate em torno de possíveis soluções ao paradoxo ganhou grandes proporções, com desenvolvimentos teóricos extremamente sofisticados, com implicações teóricas que extrapolam e muito o problema em questão. Em todos esses casos explorados na Tese, as possíveis soluções ao paradoxo foram resultado de modelos formais bas-

de tais teses é condicional para meu argumento.

7 Ver Joyce (1998), Easwaran (2016) e Easwaran e Fitelson (2015).

8 Enquanto "verdade" é tomado como um conceito qualitativo, "acurácia" é uma noção quantitativa: um estado doxástico pode ser mais ou menos acurado de acordo com sua proximidade com a verdade.

9 O princípio de dominância garante que, se A e B são duas atitudes distintas, A é dominado por B se, e somente se, para qualquer mundo possível em W, a utilidade de B é maior que a utilidade de A.

tante complexos com agentes cognitivamente perfeitos. Mas isso não significa que as abordagens apresentadas durante a Tese não sejam boas soluções simplesmente por que agentes reais não são capazes de cumprir os requisitos racionais de tais modelos. Mais do que isso, nesses casos, os modelos podem precisamente explicar o que há de errado com nossas suposições epistêmicas. Agentes idealizados são apenas recursos teóricos para aperfeiçoar nosso entendimento sobre maximização de verdade em conjuntos de crenças, não exemplos de seres humanos à ser seguido. Desse modo, o papel da modelagem epistêmica em problemas filosóficos sobre loterias é aperfeiçoar nossa compreensão sobre a relação entre crença racional e probabilidade subjetiva no que diz respeito a maximização de verdade, não sobre como devemos apostar em loterias reais.

O que quero dizer com tudo isso é que quando alguém não satisfaz um determinado requisito racional isso não implica necessariamente que ele quebrou uma regra oriunda de nossas intuições como “críticos do raciocínio”, nem significa que seu comportamento deixou de seguir algum padrão descritivo, mas talvez ele simplesmente não maximizou verdade (ou acurácia) em seu conjunto de crenças. E aqui podemos voltar ao Dilema de Harman e adicionar uma nova opção ao pacote de escolhas teóricas possíveis: o entendimento da racionalidade epistêmica pode ser feito considerando como as pessoas de fato raciocinam, considerando o que é intuitivamente exigido pela racionalidade, e além disso, também pode ser feito a partir de uma observação sobre o que os modelos recomendam. Na medida em que modelos formais se enquadram bem em uma noção instrumental da racionalidade, a vantagem agora é que não precisamos atribuir tanto peso às intuições filosóficas sobre conceitos normativos em nossa busca por uma teoria da racionalidade.

Uma das agradáveis consequências da visão acima é a existência de uma clara distinção entre o que é estritamente normativo e o que é racional. Apesar de uma teoria da racionalidade ter de lidar com elementos normativos em algum momento, isso não implica que uma teoria da racionalidade nos fornece razões sobre o que devemos fazer de modo geral. O que seres humanos devem fazer com seus estados doxásticos racionais, ou seja, a pergunta “Por que ser racional”, não é uma questão própria da teoria da racionalidade, mas sim uma questão sob o domínio do campo filosófico preocupado com “razões normativas”[10] Nesse sentido, assumir que o bayesianismo é o modelo adequado para a racionalidade das crenças graduais não implica que agente bayesianos “devam” raciocinar de acordo com os axiomas da teoria da probabilidade, ou atribuir confiança máxima a todas verdades lógicas. Resumindo, se o agente epistêmico deve ou não maximizar verdade em seu sistema doxástico não é uma questão sob o domínio da racionalidade. Embora eu reconheça que esta não seja uma “solução” para problemas como o da onisciência lógica, acredito que uma melhor

[10] Sobre a distinção entre racionalidade e normatividade, ver Broome (2013) e Kolodny (2005).

compreensão da relação entre lógica, modelos, racionalidade e normatividade pode sim mitigar seu poder destrutivo.

Espero que agora esteja claro ao leitor como a atividade de modelagem epistêmica pode auxiliar nossa busca por uma teoria da racionalidade. Se a resposta para a questão “Por que devemos satisfazer requisitos racionais bayesianos?” era “Porque eles estão de acordo com nossos princípios racionais intuitivos”, agora a resposta pode ser “Porque eles maximizam verdade (acurácia) nos estados doxásticos”. Portanto, não precisamos repousar toda nossa teoria sobre intuições normativas sobre o que é exigido pela racionalidade, e acredito que isso, por si só, já é uma razão para continuar fazendo modelagem epistêmica. Por que limitar nossas opções?

APÊNDICE B – DEMONSTRAÇÃO DOS TEOREMAS PRINCIPAIS

Este Apêndice é dedicado à demonstração dos principais Teoremas utilizados ao longo da Tese. As demonstrações serão apresentadas visando facilitar a compreensão dos resultados para o leitor não familiarizado, de um modo mais simplificado do que o apresentado originalmente na literatura.

CAPÍTULO 2

Teorema 2.3.1: Seja $\mathfrak{a}$ a medida de acurácia e P uma distribuição de probabilidade subjetiva sobre W. Para cada $A \in \mathcal{A}$, e cada estado de doxástico qualitativo atribuído pela função b, temos que:

1. $Bel(A)$ maximiza acurácia esperada *sse* $P(A) \geq \frac{-F}{T-F}$
2. $Des(A)$ maximiza acurácia esperada *sse* $P(A) \leq \frac{T}{T-F}$
3. $Sus(A)$ maximiza acurácia esperada *sse* $\frac{T}{T-F} < P(A) < \frac{-F}{T-F}$

Demonstração. A demonstração desse Teorema consiste na demonstração de cada uma das três afirmações relevantes, uma para cada possível estado doxástico atribuído pela função b.

Comecemos então pela primeira afirmação de equivalência:

$$Bel(A) \text{ maximiza acurácia esperada } sse \ P(A) \geq \frac{-F}{T-F}.$$

O primeiro passo requer observar a definição de Acurácia Esperada (apresentada em 11), que diz que a acurácia esperada de $Bel(A)$ é igual a soma dos produtos $(P(w_i) \cdot \mathfrak{a}_{w_i}(Bel(A))$, em cada mundos $w_i \in W$. Uma vez que $\mathfrak{a}_W(Bel(A))$ é definida como aparece em 10, temos que $Bel(A)$ tem valor de utilidade epistêmica positivo nos mundos em que A é verdadeira e negativo quando A é falsa, representados respectivamente pelas constantes T e F. Logo, a Acurácia Esperada de $Bel(A)$ será:

$$P(A) \cdot T + P(\neg A) \cdot F.$$

Ainda, uma vez que a Medida de Acurácia garante que a crença em uma proposição verdadeira será mais acurada que a suspensão do juízo, precisamos calcular a utilidade esperada de $Bel(A)$ e ver se seu valor será sempre maior que a utilidade da suspensão do juízo, que é 0. Como nossa informação sobre A é incompleta, precisamos calcular sua acurácia esperada. Assim, dada uma proposição A qualquer, dizemos que Bel é a atitude que maximiza a acurácia esperada para A se, e somente se, $T \cdot P(A) + F \cdot P(\neg A) \geq 0$. Uma vez que $P(\neg A) = 1 - P(A)$, temos que, $Bel(A)$ maximiza a acurácia esperada *sse*

$$\begin{aligned}
T \cdot P(A) + F \cdot (1 - P(A)) &\geq 0\\
T \cdot P(A) + F - F \cdot P(A) &\geq 0\\
T \cdot P(A) - F \cdot P(A) &\geq -F\\
P(A)(T - F) &\geq -F\\
P(A) &\geq \frac{-F}{T-F}
\end{aligned}$$

Logo, mostramos que (1) é satisfeito.

Vamos agora para a segunda equivalência:

$$Des(A) \text{ maximiza acurácia esperada } sse\ P(A) \leq \frac{T}{T-F}.$$

Para demonstrar (2), vamos mostrar primeiramente que $Des(A)$ maximiza utilidade esperada se, e somente se, $Bel(\neg A)$. Esse será um passo necessário para chegar ao resultado esperado.

Sabemos que $Des(A)$ recebe valor de utilidade epistêmica T quando $w \in \neg A$, e F quando $w \in A$. Assim, dado que estamos interessados apenas na proposição A, W é dividido apenas entre as partições w_i e w_j, e a utilidade esperada de $Des(A)$ é $T \cdot P(\neg A) + F \cdot P(A)$. Do mesmo modo que fizemos na demonstração anterior, $Des(A)$ maximiza utilidade esperada quando

$$T \cdot P(\neg A) + F \cdot P(A) \geq 0.$$

E uma vez que $P(A) = 1 - P(\neg A)$, temos que

$$T \cdot P(\neg A) + F \cdot P(A) = T \cdot P(\neg A) + F \cdot (1 - P(\neg A).$$

Seguindo exatamente os passos da demonstração do item (1), podemos concluir que $Des(A)$ maximiza utilidade esperada quando $P(\neg A) \geq \frac{-F}{T-F}$, ou seja, quando $Bel(\neg A)$ maximiza utilidade esperada, mostrando assim o que queríamos inicialmente.

Agora é fácil mostrar que (2) é o caso. Uma vez que $P(\neg A) = 1 - P(A)$, temos que $Bel(\neg A)$ maximiza utilidade esperada se, e somente se:

$$\begin{aligned}
1 - P(A) &\geq \frac{-F}{T-F}\\
(T - F) \cdot (1 - P(A)) &\geq -F\\
1 \cdot (T - F) - P(A) \cdot (T - F) &\geq -F\\
T - F &\geq -F + P(A) \cdot (T - F)\\
T - F + F &\geq P(A) \cdot (T - F)\\
\frac{T}{T-F} &\geq P(A)
\end{aligned}$$

Logo, *Des*(*A*) maximiza utilidade esperada *sse* $P(A) \leq \frac{T}{T-F}$.

Finalmente, resta apenas mostrar que *Sus*(*A*) maximiza acurácia esperada *sse*

$$\frac{T}{T-F} < P(A) < \frac{-F}{T-F}.$$

Para mostrar que (3) é o caso, vamos partir da ideia de que *Sus*(*A*) maximiza a utilidade esperada quando ambos *Bel*(*A*) e *Des*(*A*) *não* maximizam utilidade esperada para uma determinada proposição *A*. A partir desse fato a demonstração é bastante simples.

Se *Bel*(*A*) não maximiza a utilidade esperada, então, pelos resultados anteriores, sabemos que $P(A) < \frac{-F}{T-F}$. Por outro lado, se *Des*(*A*) não maximiza a utilidade esperada temos que $P(A) > \frac{T}{T-F}$. Portanto, *S*(*A*) maximiza a utilidade esperada quando $P(A)$ encontra-se no intervalo $(\frac{-F}{F-T}, \frac{T}{T-F})$, ou seja quando $\frac{T}{F-T} < P(A) < \frac{-F}{T-F}$, chegando ao resultado desejado. ■

CAPÍTULO 4

Abaixo apresento a demonstração dos dois principais Teoremas sobre o Sistema de Esferas de Grove. Para isso, utilizaremos algumas propriedades sobre a relação entre intersecções e conjunto de mundos que foram omitidas (pelo bem da fluidez do material) no texto principal. Tais propriedades aparecem em Grove (1988, p. 158):

LEMA B.0.1. *Seja Γ um conjunto consistente de sentenças em $\mathcal{L}$, $[\Gamma]$ o conjunto de mundos $w_i \in W_{\mathcal{L}}$ tal que $w_i \models \Gamma$, $\bigcap[\alpha]$ a intersecção dos mundos em $[\alpha]$, com α uma sentença qualquer, e $S_i \subseteq W_{\mathcal{L}}$. As seguinte propriedades se seguem:*

i) $\bigcap[\Gamma] = \Gamma$*, para quaisquer conjuntos de sentenças* $\Gamma \subseteq \mathcal{L}$*, se a lógica subjacente for compacta.*

ii) $\bigcap S_i \neq Cn(\bot)$ *se, e somente se,* $S_i \neq \emptyset$*.*

iii) Para todo $\alpha \in \mathcal{L}$ *e* $S_i \subseteq W_{\mathcal{L}}$, $\bigcap(S_i \cap [\alpha]) = Cn(\bigcap S \cup \{\alpha\})$*.*

iv) Para todo $S_i, S_j \subseteq W_{\mathcal{L}}$*, se* $S_i \subseteq S_j$*, então* $\bigcap S_j \subseteq \bigcap S_i$*.*

v) Para todo conjunto de sentenças $\Gamma, \Delta \subseteq \mathcal{L}$*, se* $\Gamma \subseteq \Delta$*, então* $[\Delta] \subseteq [\Gamma]$*.*

Agora podemos ir aos Teoremas:

Teorema 4.2.1: Seja $\mathbb{S}$ um sistema de esferas centrado em $[\mathbf{B}]$, para um conjunto de crenças $\mathbf{B}$. Se, para cada sentença $\alpha \in \mathcal{L}$, $\mathbf{B}^*_\alpha$ é definido como $\bigcap c(\alpha)$, então o resultado da função de revisão no sistema de esferas irá satisfazer os axiomas $(\mathbf{B}^*1)$ – $(\mathbf{B}^*8)$ da teoria AGM.

Demonstração. Para este Teorema, iremos demonstrar que o conjunto de sentenças em $\bigcap c(\alpha)$, que forma nosso conjunto de crenças, satisfaz todos os axiomas de revisão da teoria AGM (Seção 4.2.1). Esse resultado foi originalmente apresentado por Grove (1988, p. 161), e o que veremos a seguir é uma simplificação da demonstração original adaptada ao presente propósito.

Conforme o Teorema, cada processo de revisão AGM será aqui representado por $\bigcap c(\alpha)$, para uma sentença α qualquer. Relembrando, temos que $c(\alpha) := [\alpha] \cap S_\alpha$, onde S_α é a menor esfera que intersecciona $[\alpha]$. Assim, $\bigcap c(\alpha)$ irá selecionar todos os mundos pertencentes a $([\alpha] \cap S_\alpha)$ e pegar apenas os elementos (sentenças) compartilhados por todos eles, o que nos dará o conjunto relevante de sentenças. Vamos então aos 8 axiomas de revisão AGM:

$(\mathbf{B}^*1)$: Nesse caso basta mostrar que, para qualquer $\alpha \in \mathcal{L}$, $\bigcap c(\alpha)$ é um conjunto de crenças. Se $[\alpha] \neq \emptyset$, existe um menor núcleo S_α que intersecciona $[\alpha]$, e nesse caso $\bigcap([\alpha] \cap S_\alpha)$ é um conjunto de crenças. Se α é uma sentença contraditória, isto é, $[\alpha] = \emptyset$, então $c(\alpha) = \emptyset$ e consequentemente $\bigcap c(\alpha)$ é um conjunto de crenças vazio.

($\mathbf{B}^*2$): Aqui devemos mostrar que $\alpha \in \bigcap c(\alpha)$. Pelo item (iii) do Lema apresentado acima, sabemos que $\bigcap(S \cap [\alpha]) = Cn(\bigcap S \cup \{\alpha\})$. Portanto, uma vez que $\alpha \in Cn((\bigcap S_\alpha) \cup \{\alpha\})$, temos que $\alpha \in \bigcap(S_\alpha \cap [\alpha]) = \bigcap c(\alpha)$, chegando assim ao resultado desejado.

($\mathbf{B}^*3$): Aqui precisamos mostrar que, para um determinado conjunto de crenças $\mathbf{B}$, temos que $\bigcap c(\alpha) \subseteq Cn(\mathbf{B} \cup \{\alpha\})$. Dado que $[\mathbf{B}]$ é a menor esfera em $\mathbb{S}$, então também é a menor esfera que intersecciona $[\alpha]$. Logo, $S_\alpha = [\mathbf{B}]$. Como $c(\alpha) = [\alpha] \cap S_\alpha$, consequentemente, $c(\alpha) = [\alpha] \cap [\mathbf{B}]$. Novamente pelo item (iii) do Lema, sabemos que $\bigcap([\alpha] \cap [\mathbf{B}])$ é logicamente equivalente a $Cn(\bigcap[\mathbf{B}] \cup \{\alpha\})$. Pelo item (i) do Lema, $\bigcap[\mathbf{B}] = \mathbf{B}$. Portanto, podemos concluir que todo elemento de $\bigcap c(\alpha)$ é elemento de $Cn(\mathbf{B} \cup \{\alpha\})$, ou seja, $\bigcap c(\alpha) \subseteq Cn(\mathbf{B} \cup \{\alpha\})$.

($\mathbf{B}^*4$): Segue-se exatamente pela mesma demonstração acima.

($\mathbf{B}^*5$): Aqui demonstraremos uma versão análoga do axioma ($\mathbf{B}^*5$), que diz que se α não é uma sentença contraditória, então a revisão de $\mathbf{B}$ por α será também consistente, isto é, $\bigcap c(\alpha) \neq \emptyset$. Primeiro vamos mostrar que se $[\alpha] \neq \emptyset$, então $\bigcap c(\alpha)$ é consistente. Dado que $[\alpha] \neq \emptyset$, então existe uma menor esfera S_α que intersecciona $[\alpha]$. Logo, $[\alpha] \cap S_\alpha = c(\alpha) \neq \emptyset$, e pelo item (ii) do Lema, $\bigcap c(\alpha)$ é consistente. Agora, se $\bigcap c(\alpha)$ é consistente, então $\bigcap([\alpha] \cap S_\alpha) \neq \emptyset$, logo $[\alpha]$ não pode ser um conjunto vazio.

($\mathbf{B}^*6$): Aqui devemos mostrar que, se $\vdash \alpha \leftrightarrow \beta$, então $\bigcap c(\alpha) = \bigcap c(\beta)$. Se α é logicamente equivalente a β, então $[\alpha] = [\beta]$, isto é, ambas as sentenças são verdadeiras no mesmo conjunto de mundos possíveis. Consequentemente, para um determinado conjunto de crenças $\mathbf{B}$, temos que $[\alpha] \cap \mathbf{B} = [\beta] \cap \mathbf{B}$, o que significa que $c(\alpha) = c(\beta)$ e consequentemente $\bigcap c(\alpha) = \bigcap c(\beta)$.

($\mathbf{B}^*7$): Devemos demonstrar que, para um determinado conjunto de crenças $\mathbf{B}$, temos que $\bigcap c(\alpha \wedge \beta) \subseteq Cn((\bigcap c(\alpha)) \cup \{\beta\})$. Como sabemos, $\alpha \wedge \beta \rightarrow \alpha$ é uma verdade lógica. Assim, qualquer esfera que intersecciona $[\alpha \wedge \beta]$ também intersecciona $[\alpha]$. Com isso, temos que

$$S_\alpha \subseteq S_{\alpha \wedge \beta}$$
$$[\alpha] \cap [\beta] \cap S_\alpha \subseteq [\alpha] \cap [\beta] \cap S_{\alpha \wedge \beta}.$$

Uma vez que $[\alpha] \cap [\beta] = [Cn(\{\alpha\} \cup \{\beta\})] = [\alpha \wedge \beta]$, chegamos em:

$$[\alpha] \cap [\beta] \cap S_\alpha \subseteq [\alpha \wedge \beta] \cap S_{\alpha \wedge \beta}.$$

Para finalizar, precisamos retomar o item (iv) do Lema B.0.1, que diz que, se $S_i \subseteq S_j$, então $\bigcap S_j \subseteq \bigcap S_i$. Como todos os membros da relação acima são conjuntos de mundos possíveis, temos que

$$\bigcap([\alpha \wedge \beta] \cap S_{\alpha \wedge \beta}) \subseteq \bigcap([\alpha] \cap [\beta] \cap S_\alpha).$$

Uma vez que $[\alpha] \cap S_\alpha = c(\alpha)$, temos que

$$\bigcap c(\alpha \wedge \beta) \subseteq \bigcap([\alpha] \cap [\beta] \cap S_\alpha).$$

E como $c(\alpha) = [\alpha] \cap S_\alpha$, pelo item (iii) do Lema B.0.1, temos que

$$\bigcap c(\alpha \wedge \beta) \subseteq \bigcap(c(\alpha) \cap [\beta]) = Cn((\bigcap c(\alpha)) \cup \{\beta\})$$

Chegando assim ao resultado desejado.

($\mathbf{B}^*8$): Finalmente, no último axioma precisamos mostrar que, se $\neg\beta \notin \bigcap c(\alpha)$, então $Cn((\bigcap c(\alpha)) \cup \{\beta\}) \subseteq \bigcap c(\alpha \wedge \beta)$. Se $\neg\beta \notin \bigcap c(\alpha)$, isso significa que $[\beta] \cap [\alpha] \cap S_\alpha \neq \emptyset$ e consequentemente $[\alpha \wedge \beta] \cap S_\alpha \neq \emptyset$. Desse modo, existe ao menos um mundo onde q é verdadeira e esse mundo pertence a menor esfera que intersecciona α. Logo, a menor esfera que intersecciona $\alpha \wedge \beta$ será um subconjunto da menor esfera que intersecciona apenas α, ou seja, $S_{\alpha\wedge\beta} \subseteq S_\alpha$. Argumentando de modo semelhante à demonstração anterior, chegamos que $Cn(\bigcap c(\alpha) \cup \{\beta\}) \subseteq \bigcap c(\alpha \wedge \beta)$. ■

Teorema 4.2.2: Seja $* : \mathbf{B} \times \mathcal{L} \longrightarrow \mathbf{B}$ uma função qualquer que satisfaça ($\mathbf{B}^*1$) – ($\mathbf{B}^*8$). Então, para qualquer conjunto (fixo) de crenças **B**, existe um sistema de esferas $\mathbb{S}$ que é centrado em [**B**] e que satisfaz $\mathbf{B}^*_\alpha = \bigcap c(\alpha)$ para qualquer $\alpha \in \mathcal{L}$.

Demonstração. Aqui utilizaremos as propriedades de um sistema de esferas que foram apresentadas no texto na Definição 4.2.1.3. Para facilitar nosso trabalho, reproduzo-as novamente aqui:

($\mathbb{S}1$) $\mathbb{S}$ é totalmente ordenado por $\subseteq$; ou seja, se S_1 e S_2 são elementos de $\mathbb{S}$, então $S_1 \subseteq S_2$ ou $S_2 \subseteq S_1$.

($\mathbb{S}2$) [**B**] é o $\subseteq$ –mínimo de $\mathbb{S}$; ou seja, dado $[\mathbf{B}] \in \mathbb{S}$, se $S_1 \in \mathbb{S}$ então $[\mathbf{B}] \subseteq S_1$.

($\mathbb{S}3$) $W_\mathcal{L} \in \mathbb{S}$, e é o maior elemento de $\mathbb{S}$.

($\mathbb{S}4$) Seja $\alpha \in \mathcal{L}$ uma sentença e S uma esfera qualquer em $\mathbb{S}$ que intersecciona $[\alpha]$, então existe uma menor esfera em $\mathbb{S}$ que intersecciona $[\alpha]$; ou seja, existe uma menor esfera S_i tal que $S_i \cap [\alpha] \neq \emptyset$, e para todo $S_j \in \mathbb{S}$, se $S_j \cap [\alpha] \neq \emptyset$ então $S_i \subseteq S_j$.

Agora podemos avançar em direção à demonstração do Teorema. Nosso objetivo aqui será construir um sistema de esferas $\mathbb{S}$ que seja centrado em [**B**], e cujo processo de revisão por uma sentença seja, de fato, um procedimento sobre $\mathbb{S}$ de tal modo que $[\mathbf{B}^*_\alpha] = [\alpha] \cap S_\alpha = c(\alpha)$.

Seja S' a classe de todas esferas S_i em $W_\mathcal{L}$ que satisfazem as seguinte condições:

i) $\forall w_i \in S_i, \exists \alpha \in \mathcal{L} : w_i \in [\mathbf{B}^*_\alpha]$.

ii) Se $[\alpha] \cap S_i \neq \emptyset$ para qualquer $\alpha \in \mathcal{L}$, então $[\mathbf{B}^*_\alpha] \subseteq S_i$.

Seja $\mathbb{S} = S' \cup W_{\mathcal{L}}$. Precisamos mostrar, primeiramente, que $\mathbb{S}$ é um sistema de esferas centrado em **B**, isto é, que $S' \cup W_{\mathcal{L}}$ satisfaz as condições $\mathbb{S}1-4$ da definição acima. As demonstrações aqui são bem diretas: por definição, $S' \cup W_{\mathcal{L}}$ é totalmente ordenado por $\subseteq$; pela condição (ii), podemos afirmar que $[\mathbf{B}]$ é a menor esfera em $\mathbb{S}$; também temos, por definição, que $W_{\mathcal{L}}$ é a maior esfera em $\mathbb{S}$; e a condição (i) garante a existência de uma menor esfera S_i tal que $S_i \cap [\alpha] \neq \emptyset$.

Agora vamos à segunda parte da demonstração: mostrar que um processo de revisão em **B** é resultado de um procedimento em $\mathbb{S}$ de tal modo que $[\mathbf{B}^*_\alpha] = [\alpha] \cap S_\alpha$.

Primeiramente, se $\vdash \neg\alpha$, então, pelo axioma ($\mathbf{B}^*5$), $\mathbf{B}^*_\alpha$ é o conjunto de crenças contraditório. Então vamos supor que $\nvdash \neg\alpha$. Desse modo, pelas propriedades ($\mathbb{S}2$) e ($\mathbb{S}4$), temos que

$$[\mathbf{B}^*_\alpha] \subseteq [\alpha] \cap (\bigcap\{S_i \in \mathbb{S} : S_i \cap [\alpha] \neq \emptyset\}).$$

Agora basta mostrar que a relação de subconjunto acima é também uma relação de igualdade, e encontrar uma esfera S_i tal que $S_i = S_\alpha$, e assim estará finalizada a demonstração.

Começando, como precisamos demonstrar uma afirmação de existência, basta encontrar um S_i que satisfaça as condições desejadas, mais precisamente, as condições (i) e (ii) apresentadas acima. Desse modo, seja S_i igual a união de todos $[\mathbf{B}^*_\beta]$ tal que $[\alpha] \subseteq [\beta]$, com $\beta \in \mathcal{L}$. Mais precisamente:

$$S_i := \bigcup\{[\mathbf{B}^*_\beta] : [\alpha] \subseteq [\beta]\}$$

Com isso, temos que S_i é uma esfera, e a primeira condição (i) é satisfeita automaticamente, uma vez que existe um β tal que $w \in [\mathbf{B}^*_\beta]$, para todo $w \in S_i$.

A segunda condição (ii) é um pouco mais complexa, e demandará uma maior atenção do leitor. Suponha que $[\gamma] \cap S_i \neq \emptyset$, para $\gamma \in \mathcal{L}$. Isso significa que $[\gamma] \cap \mathbf{B}^*_\beta \neq \emptyset$, para uma sentença β tal que $[\alpha] \subseteq [\beta]$. Sabemos que $[\alpha] \subseteq [\beta] \subseteq [\beta \vee \gamma]$ e, consequentemente, pela definição de S_i,

$$[\mathbf{B}^*_{\beta\vee\gamma}] \subseteq S_i. \tag{18}$$

Agora vamos supor, por redução ao absurdo, que $\neg\gamma \in \mathbf{B}^*_{\beta\vee\gamma}$. Sendo assim, $\neg\beta \notin \mathbf{B}^*_{\beta\vee\gamma}$. logo, pelos axiomas ($\mathbf{B}^*7$) e ($\mathbf{B}^*8$), temos que $\mathbf{B}^*_{(\beta\vee\gamma)\wedge\beta} = Cn(\bigcap c(\beta \vee \gamma) \cup \{\beta\})$, o que nos mostra que $\mathbf{B}^*_\beta = \mathbf{B}^*_{\beta\vee\gamma}$, e portanto, $\neg\gamma \in \mathbf{B}^*_\beta$. Mas esta última afirmação contradiz nossa suposição inicial que diz que $[\gamma] \cap \mathbf{B}^*_\beta \neq \emptyset$. Logo, devemos rejeitar a afirmação de que $\neg\gamma \in \mathbf{B}^*_{\beta\vee\gamma}$, fechando assim nossa hipótese para a redução ao absurdo.

Dando continuidade, uma vez que $\neg\gamma \notin \mathbf{B}^*_{\beta\vee\gamma}$, pelos axiomas ($\mathbf{B}^*7$) e ($\mathbf{B}^*8$), temos que $\mathbf{B}^*_{(\beta\vee\gamma)\wedge\beta} = Cn(\bigcap c(\beta \vee \gamma) \cup \{\beta\})$, o que garante que $\mathbf{B}^*_{\beta} = \mathbf{B}^*_{\beta\vee\gamma}$. Portanto, $\neg\gamma \in \mathbf{B}^*_{\beta}$, o que contradiz nossa suposição inicial de que $[\gamma] \cap \mathbf{B}^*_{\beta} \neq \emptyset$.

Uma vez que $\neg\gamma \notin \mathbf{B}^*_{\beta\vee\gamma}$, então, pelo axioma ($\mathbf{B}^*8$), temos que

$$Cn(\bigcap c(\beta \vee \gamma) \cup \{\gamma\}) \subseteq \mathbf{B}^*_{(\beta\vee\gamma)\wedge\gamma} = \mathbf{B}^*_{\gamma}.$$

Uma vez que $(\beta \vee \gamma) \in Cn(\{\gamma\})$, pela condição (iii) do Lema B.0.1, temos que

$$Cn(\bigcap c(\beta \vee \gamma) \cup \{\gamma\}) = \bigcap(c(\beta \vee \gamma) \cap [\beta \vee \gamma]) = \mathbf{B}^*_{\beta\vee\gamma}.$$

E pela condição (v) do Lema B.0.1, e pela equação 18,

$$[\mathbf{B}^*_{\gamma}] \subseteq [\mathbf{B}^*_{\beta\vee\gamma}] \subseteq S_i.$$

Logo, a segunda condição é satisfeita.

Por fim, se $[\alpha] \cap [\mathbf{B}^*_{\beta}] \neq \emptyset$, então $\neg\alpha \notin \mathbf{B}^*_{\beta}$, e portanto, pelos axiomas ($\mathbf{B}^*7$) e ($\mathbf{B}^*8$), temos que $\mathbf{B}^*_{\alpha\wedge\beta} = Cn(\cap c(\beta) \cup \{\alpha\})$, ou seja, pela condição (iii) do Lema B.0.1, $[\alpha] \cap [\mathbf{B}^*_{\beta}] = [\mathbf{B}^*_{\alpha\wedge\beta}]$, e se $[\alpha] \subseteq [\beta]$, então $[\mathbf{B}^*_{\alpha\wedge\beta}] = [\mathbf{B}^*_{\alpha}]$. Logo, temos que $[\alpha] \cap S_i = \bigcup\{[\alpha] \cap [\mathbf{B}^*_{\beta}] : [\alpha] \subseteq [\beta]\}$, onde ao menos um dos membros de $[\alpha] \cap [\mathbf{B}^*_{\beta}]$ é $[\mathbf{B}^*_{\alpha}]$. Sendo assim, encontramos nosso S_i tal que $[\alpha] \cap S_i = [\mathbf{B}^*_{\alpha}]$. Portanto, $\mathbb{S}$ satisfaz $[\mathbf{B}^*_{\alpha}] = [\alpha] \cap S_{\alpha}$, e consequentemente $\mathbf{B}^*_{\alpha} = \cap c(\alpha)$ que era precisamente o que queríamos. ■

Abaixo apresento, de forma resumida, demonstração dos dois principais resultado de Lin e Kelly sobre a Teoria Probalógica. O primeiro deles irá consiste em um teorema sobre a incompatibilidade entre a teoria AGM e o rastreamento bayesiano. O segundo deles consiste em um teorema da representação sobre a relação entre a Teoria Probalógica e o método de revisão de crença não-monotônico baseado em uma ordem de plausibilidade ao estilo Shoham (1987). Para facilitar, reproduzo novamente aqui as definições que serão fundamentais para os resultados a seguir.

Dizemos que um PP é *sensível* se satisfaz as seguintes condições:

1. ser não-cético (ou seja, permite crença com probabilidade menor que 1);
2. não-opinado (ou seja, permite crenças incompletas ou disjuntivas);
3. consistente (isto é, não existe uma distribuição de probabilidade P tal que $Bel_P(\bot)$); e
4. monótono de canto (isto é, se $Bel_P(A)$ e temos que $P(A \mid E) > P(A)$, então $Bel_{P|E}(A)$).

Dizemos que um PP *rastreia condicionalização bayesiana* se, e somente se, para todo $A_i \in \mathcal{A}$, $P(E) > 0$, e toda distribuição $P \in \mathcal{P}$, temos que:

$$Bel_P(A_i \mid E) \text{ sse } Bel_{P|E}(A_i).$$

Por fim, dizemos que um PP é *cumulativo* quando possui como método de revisão de crença uma teoria que satisfaz Monotonicidade Racional:

$$\text{Se } A \mid\!\sim B \text{ e } A \not\mid\!\sim \neg C, \text{ então } A \wedge C \mid\!\sim B$$

ou simplesmente, os axiomas $\mathbf{B}^*3$ e $\mathbf{B}^*4$ da teoria AGM:

($\mathbf{B}^*3$) $\mathbf{B}^*_A \subseteq \mathbf{B}^+_A$;

($\mathbf{B}^*4$) Se $\neg A \notin \mathbf{B}$, então $\mathbf{B}^+_A \subseteq \mathbf{B}^*_A$;

Vamos então aos Teoremas.

Teorema 4.3.1: Se $\mathcal{Q}$ é uma questão com pelo menos 3 respostas, nenhum PP sensível que rastreia condicionalização bayesiana é cumulativo.

Demonstração. Munido das definições acima, podemos agora ir diretamente à demonstração. O que será apresentado a seguir consiste em uma versão simplificada do resultado apresentado originalmente pelos autores em Lin e Kelly (2012a,b). A proposta é apresentar a intuição subjacente à demonstração, da forma mais acessível possível, deixando de lado alguns detalhes técnicos da demonstração original.

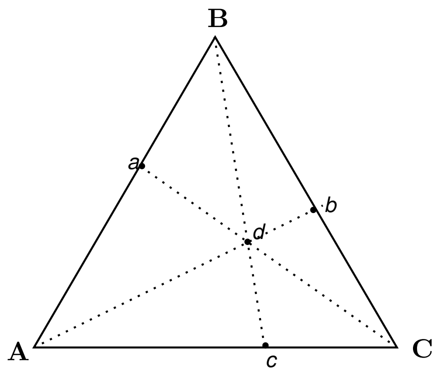

Figura 16 – Distribuições de probabilidade arbitrárias nos pontos a, b, c e d.

Considere nossa questão relevante $\mathcal{Q}$ como uma partição das proposições A, B e C, e $\mathcal{P}$ o conjunto de todas possíveis distribuições de probabilidade sobre os elementos de $\mathcal{Q}$, conforme viemos trabalhando até aqui. A Figura 16 reproduz novamente esses elementos centrais, e adiciona alguns outros que serão relevantes em nossa demonstração (na figura, os vértice do triângulo são os pontos onde a distribuição de probabilidade é máxima, isto é, igual a 1, para cada uma das três respostas disponíveis; As arestas do triângulo representam as regiões onde respostas disjuntivas possuem probabilidade máxima). Relembrando, quando $\mathcal{Q}$ possui apenas duas respostas, $\mathcal{P}$ pode ser entendido geometricamente como o conjunto de todos os pontos pertencentes ao intervalo $[0, 1]$, ou de modo equivalente, a uma das arestas de medida unitária no triângulo. Quando $\mathcal{Q}$ possui três respostas, $\mathcal{P}$ é o conjuntos dos pontos no interior do $2-$*simplex*, como vimos ao longo dos capítulos 3 e 4. Em Lin e Kelly (2012b, p. 973), os autores mostram como o teorema acima é válido para qualquer questão $\mathcal{Q}$ com pelo menos três respostas. Por razão de simplicidade, a demonstração a seguir será limitada ao caso onde $\mathcal{Q}$ possui *exatamente* três respostas.

Uma vez que trabalharemos em uma partição com apenas 3 repostas possíveis, os recursos geométricos utilizados aqui estarão limitados a $\mathbb{R}^1$ e $\mathbb{R}^2$, o que facilita bastante nosso trabalho. A demonstração abaixo se baseia fundamentalmente em uma concepção geométrica do problema, e tentarei apresentar os resultados do modo mais acessível possível para que o leitor possa ter uma ideia geral dos passos relevantes para a demonstração do resultado. Um conceito matemático que será importante é a noção de "vizinhança aberta". Resumidamente, em $\mathbb{R}^1$, como temos no intervalo $[0, 1]$, uma vizinhança aberta V de um ponto c é um intervalo aberto contendo c em seu centro, sendo c um ponto qualquer em $(0, 1)$. Em $\mathbb{R}^2$, uma vizinhança aberta de um ponto c é um círculo de raio r com centro em c, tal que $r > 0$. A noção de vizinhança aberta, como veremos abaixo, será importante em alguns dos passos necessários na demonstração.

Como ficará claro a seguir, o resultado relevantes será encontrado por meio de uma série de passos. Mas, basicamente, a demonstração pressupõe um PP que satisfaz todas as condições do Teorema, mas que termina sendo um PP opinado, o que entra em contradição com a exigência (2) de um PP sensível. Para isso, faremos referência diversas vezes à Figura 16.

Primeiramente, considere $\mathcal{Q}$ nossa questão relevantes, e $\mathcal{P}$ uma distribuição qualquer sobre as três resposta possíveis. Isso significa que $\mathcal{P}$ é um ponto qualquer em nosso $2-$*simplex*. Agora seja V uma vizinhança aberta no ponto $\mathcal{P}$. Pela condição de não-ceticismo, existe um V não vazio onde qualquer $P_i \in V$ implica que a proposição A é aceita como reposta mais forte. Relembrando, um PP é dito não-cético se, e somente se, para cada resposta $A \in \mathcal{Q}$ existe uma distribuição $P_i \in \mathcal{P}$ tal que $P_i(A) < 1$ e $Bel_{P_i}(A)$, segundo o PP em questão. Uma vez que $V \neq \emptyset$, existe uma distribuição P_i em V que atribui probabilidade estritamente menor que 1 para alguma resposta em $\mathcal{Q}$.

A partir disso, nosso objetivo agora será mostrar que para quaisquer arestas $\overline{A_iA_j}$ do triângulo, $\mathcal{P}$ contém no máximo um ponto onde o PP relevante implica que $Bel(A_i \vee A_j)$. Mais precisamente, demonstraremos a seguinte afirmação:

PROPOSIÇÃO B.0.2. *Para todo ponto $P_j \in \overline{A_iA_j}$, se $Bel_{P_i}(A_i \vee A_j)$, então $P_i = P_j$.*

Considere $P_i \in \mathcal{P}$ um ponto qualquer no interior do triângulo $\triangle Aad$, conforme vemos na Figura 16. Primeiramente, sabemos que a probabilidade condicional $P_i(\cdot \mid \neg C)$ está definida (para ver como isso é o caso, basta olhar a aresta $\overline{AB}$ na Figura 16, e observar que uma vizinhança aberta V representa um intervalo no segmento $\overline{AB}$, e a condição de não-ceticismo garante a probabilidade positiva de ao menos um deles, e $A \vee B = \neg C$). Uma vez que $P_i(\cdot \mid \neg C)$ existe, como, por suposição, $P_i \in \triangle Aad$, então $P_i \mid \neg C \in \overline{Aa}$. Agora seja a um ponto arbitrário em $\overline{AB}$ tal que, se $P_i = a$, então P_i é a distribuição de probabilidade com menor valor para $P(A)$ mas que garante $Bel_{P_i}(A)$ (e o mesmo raciocínio vale para os pontos b e c, respectivamente aos vértices $\overline{BC}$ e $\overline{AC}$). Logo, como $P_i \mid \neg C \in \overline{Aa}$, temos que $A \in [\mathbf{B}]_{P_i|\neg C}$.

Agora, pela condição de rastreamento, chegamos em $A \in [\mathbf{B}]_{P_i} * \neg C$, e pelo axioma ($\mathbf{B}^*3$) temos

$$[\mathbf{B}]_{P_i} \wedge \neg C \models A.$$

Isso significa que temos apenas três opções de resposta para o conjunto $[\mathbf{B}]_{P_i}$ (que por definição são mutuamente exclusivas):

$$\text{(i) } Bel_{P_i}(A); \quad \text{ou (ii) } Bel_{P_i}(C); \quad \text{ou (iii) } Bel_{P_i}(A \vee C).$$

A seguir eliminaremos as duas últimas opções, ficando apenas com a primeira.

Suponha, por redução, que $Bel_{P_i}(C)$ ou $Bel_{P_i}(A \vee C)$. Uma vez que nosso PP satisfaz ($\mathbf{B}^*4$), dado que $[\mathbf{B}]_{P_i} * \neg A$ é consistente com $\neg A$, então

$$[\mathbf{B}]_{P_i} * \neg A \models [\mathbf{B}]_{P_i} \wedge \neg A.$$

Pela condição de rastreamento, o lado esquerdo equivale a $[\mathbf{B}]_{P_i|\neg A}$, e pela hipótese para redução, o lado direito é C. Logo,

$$[\mathbf{B}]_{P_i|\neg A} \models C.$$

Mas, uma vez que $[\mathbf{B}]$ é consistente e C é uma resposta completa para $\mathcal{Q}$, $[\mathbf{B}]_{P_i|\neg A} = C$. Isso significa que $P_i \mid \neg A$ é um ponto na aresta $\overline{BC}$ no $2-$*simplex*, e $P_i \mid \neg A$ pertence a reta $\overline{Cb}$, pela definição do ponto c. Mas isso é impossível, uma vez que supomos inicialmente que P_i é um ponto em $\triangle Aad$. Portanto, eliminamos as opções (ii) e (ii) e podemos concluir que $Bel_{P_i}(A)$.

O próximo passo consiste em mostrar que A é aceita como resposta mais forte para $\mathcal{Q}$ no segmento $\overline{Ac}$. Seja $P_j \in \mathcal{P}$ um ponto arbitrário no $2-$*simplex* tal que $P_j = P_i \mid (A \vee C)$. Como vimos anteriormente, $A \in [\mathbf{B}]_{P_i}$. Uma vez que $A \vee C$ é consistente com $[\mathbf{B}]$, por $(\mathbf{B}^*4)$:

$$[\mathbf{B}]_{P_i} * (A \vee C) \models [\mathbf{B}]_{P_i} \wedge (A \vee C).$$

Sabemos, pela condição de rastreamento, que o lado esquerdo da afirmação acima é igual a $[\mathbf{B}]_{P_i|A\vee C}$ e, uma vez que $A \in [\mathbf{B}]_{P_i}$, o lado direito é simplesmente $Bel_{P_i}(A)$. Desse modo,

$$[\mathbf{B}]_{P_i|A\vee C} \models Bel_{P_i}(A).$$

Dado $[\mathbf{B}]$ consistente e A uma resposta completa, temos que $[\mathbf{B}]_{P_i|A\vee C} = A$. E como, por suposição, $P_j = P_i \mid (A \vee C)$, temos que $[\mathbf{B}]_{P_j} = A$. Com exatamente o mesmo raciocínio, podemos mostrar que C é aceita como resposta mais forte para $\mathcal{Q}$ no segmento $\overline{Cc}$. Com isso, demonstramos o que afirmava a proposição B.0.2: $\mathcal{P}$ contém no máximo um ponto onde $A \vee C$ é aceito como resposta mais forte, e isso é o caso para quaisquer arestas $\overline{A_iA_j}$ no $2-$*simplex*.

Para finalizar, basta mostrar que o resultado anterior expresso pela proposição B.0.2 entra em contradição com as hipóteses iniciais para nosso PP relevante. Vamos supor, por redução, que o PP em questão não é opinado. Um PP é dito não-opinado se, e somente se, existe uma vizinhança aberta V de $\mathcal{P}$ onde alguma resposta incompleta, disjuntiva, é aceita. Como vimos no início, existe um $P_i \in V$ que atribui probabilidade positiva para ambos A e B. Então existe uma vizinhança aberta V' de P_i tal que $V' \subseteq V$. Agora podemos mover o centro de V' e obter um novo ponto P_j, com $P_j \neq P_i$, de tal modo que $P_j \in V' \subseteq V$, e P_j atribui probabilidade positiva para ambos A e C. Segue-se que tanto $P_i \mid A \vee C$ quanto $P_j \mid A \vee C$ são distribuições be definidas e pontos distintos no $2-$*simplex*. Uma vez que P_i e P_j pertencem a V temos que $Bel_{P_i}(A \vee C)$ e $Bel_{P_j}(A \vee C)$. Então existe $P_i \neq P_j$ tal que $Bel_{P_i}(A \vee C)$ e $Bel_{P_j}(A \vee C)$, o que contradiz a Proposição B.0.2. Portanto, pela hipótese de redução ao absurdo, nosso PP é opinado, ou seja, ele não pode ser sensível, e isso contradiz as suposições iniciais. Demonstramos assim a afirmação central de nosso teorema: nenhum PP sensível que rastreia condicionalização bayesiana é cumulativo. ■

Teorema 4.3.2: A teoria Probalógica com um método de revisão de crença baseado em uma ordem de plausibilidade em um Sistema Preferencial, conforme Definição 32, cumpre os seguintes requisitos:

i é sensível (se $\mathcal{Q}$ contém ao menos três respostas completas);

ii rastreia condicionalização bayesiana; e

iii satisfaz Dedutivismo Hipotético e Raciocínio de Caso

Demonstração. Antes de qualquer coisa, vamos retomar aqui o que diz a definição 32:

> Sejam w_i e w_j respostas completas para $\mathcal{Q}$, com $P(w_j) > 0$, e t um limiar com valor estritamente maior que 1:
>
> $$w_i \prec_P w_j \ sse \ \frac{P(w_i)}{P(w_j)} > t$$

Novamente, a demonstração a seguir consiste em uma adaptação dos resultados originais apresentados em (LIN; KELLY, 2012b), e nos restringiremos para o caso com apenas 3 respostas completas. Iniciaremos mostrando que a teoria Probalógica com uma ordem de plausibilidade de um Sistema Preferencial é *sensível.*

(i)

Chamemos de $\widehat{w_A}$ o vértice do nosso $2-$*simplex* onde a proposição A é aceita como uma resposta completa para $\mathcal{Q}$ (semelhante a figura 16 que vimos na definição anterior), ou seja, onde $P(A) = 1$. Seja d a distância entre dois pontos no triângulo $\triangle ABC$ e definimos d_{ij}, com $i, j \in \{A, B, C\}$ e $i \neq j$, como:

$$d_{ij} = \inf\{ \ \|\widehat{w_i} - P\| : P \in \mathcal{P}, \frac{P(w_i)}{P(w_j)} \leq t \ \}.$$

Relembrando, dizemos que um PP é não cético quando existe uma vizinhança V não vazia onde qualquer $P_i \in V$ implica que uma proposição é aceita como mais forte em P_i. Isso significa que, para ver que nosso PP é não-cético, basta mostrar que $d_{ij} > 0$ pela definição acima de distância, uma vez que A é aceita em $\widehat{w_A}$. Para isso, por redução ao absurdo, vamos supor que $d_{ij} = 0$. Se isso for o caso, então existe uma sequência P_n de pontos, com $n \in \mathbb{N}$, tal que $\frac{P_n(w_i)}{P_n(w_j)} \leq t$, e $\lim_{n\to\infty} \|w_i - P_n\| = 0$. Isso significa que, nesse caso, $\lim_{n\to\infty} P_n(w_i) = 1$ e $\lim_{n\to\infty} P_n(w_j) = 0$. Uma vez que, pelas propriedades de limite, se $\lim_{n\to\infty} P_n(w_j) = 0$ então $\frac{1}{\lim_{n\to\infty} P_n(w_j)}$ vai ao infinito positivo, podemos concluir que, para um m suficientemente grande, temos que $\frac{P_m(w_i)}{P_m(w_j)} > t$, o que contradiz nossa definição de d_{ij}. Portanto, $d_{ij} > 0$. Finalmente, basta tomar uma vizinhança V de raio d_{ij} em torno do vértice w_A e notar que, para todo $P \in V$, $\|w_A - P\| < d_{ij}$, e isso implica que $\frac{P(A)}{P(w_j)} > t$, para qualquer resposta verdadeira em w_j, o que significa que cada resposta completa diferente de A é rejeitada pelo PP em questão, e, consequentemente, temos que nosso PP não é cético.

Para mostrar que o PP é não-opinado, basta mostrar que uma disjunção particular, digamos, $A \vee B$, é aceita em algum intervalo aberto. Considere $P' \in \mathcal{P}$ tal que $\frac{P'(A)}{P'(B)} = 1$, e $P'(C) = 0$, semelhante ao conjunto de respostas $\{A, B, C\}$ que da figura 16. Agora vamos definir o seguinte conjunto de pontos a, b, c, d em nosso $2 - simplex$ como:

$$a = inf\{\|P' - P\| : P \in \mathcal{P}, \frac{P(A)}{P(C)} \leq t\};$$
$$b = inf\{\|P' - P\| : P \in \mathcal{P}, \frac{P(B)}{P(C)} \leq t\};$$
$$c = inf\{\|P' - P\| : P \in \mathcal{P}, \frac{P(A)}{P(B)} > t\};$$
$$d = inf\{\|P' - P\| : P \in \mathcal{P}, \frac{P(B)}{P(A)} > t\};$$

Pelo mesmo procedimento que fizemos na demonstração acima de não-ceticismo, podemos concluir que $a, b > 0$. Para ver como $c, d > 0$, basta supor, por redução, que, por exemplo, $c = 0$, e por um raciocínio semelhante ao caso acima chegaremos que existe uma distribuição P_m tal que $\frac{P_m(A)}{P_m(B)} < t$, o que contradiz a construção inicial. Com isso podemos concluir que a, b, c, d são estritamente maiores que 0.

Para terminar mostrando que nosso PP é não-opinado, vamos tomar uma vizinhança aberta V' com centro em P' e de raio r^* tal que:

$$r^* = \inf\{a, b, c, d\}$$

Agora basta mostrar que $A \vee B$ é aceita como resposta mais forte em V', ou seja, para cada $P \in V'$, temos que:

$$\frac{P(A)}{P(C)} > t;$$
$$\frac{P(B)}{P(C)} > t;$$
$$\frac{P(A)}{P(B)} < t;$$
$$\frac{P(B)}{P(A)} < t.$$

duas primeiras afirmações se seguem do mesmo modo que no caso utilizado para o não-ceticismo. A terceira e a quarta afirmação se seguem do seguinte fato: para qualquer $P \in V'$, temos que $\|P' - P\| < r^*$, o que implica que $\|P' - P\| < c$ e consequentemente $\frac{P(A)}{P(B)} \leq t$, chegando assim ao resultado desejado, comprovando que nosso PP não é opinado.

A consistência segue imediatamente do fato de um Sistema Preferencial não possuir uma cadeia descendente infinita de mundos possíveis, ou seja, possui uma ordem bem fundada.

E por fim, a Monotonicidade de Canto se segue da seguinte propriedade algébrica: a cotação $\frac{P(A)}{P(B)}$ aumenta monotonicamente se a distribuição de probabilidade P se move em direção ao vértice $\widehat{w_A}$ onde A é aceito como resposta mais forte. Com isso, mostramos que nosso PP é sensível.

(ii)

Assumindo que nosso PP possui uma ordem de plausibilidade em um sistema Preferencial, basta mostrar que uma resposta é a mais plausível em $[\mathbf{B}]_P * E$ se, e somente se, ela é a mais plausível em $[\mathbf{B}]_{P|E}$. Isso ocorre recursivamente pelo seguinte procedimento: dada uma nova informação E, seja $[\mathbf{B}_{\prec}] * E$ a disjunção das repostas mais plausíveis para $\mathcal{Q}$ no que diz respeito a ordem de plausibilidade $\prec$ que são logicamente compatíveis com E. Assim, um método de revisão qualitativo é definido de tal modo que $[\mathbf{B}]_P = [\mathbf{B}_{\prec_P}]$, onde $[\mathbf{B}_{\prec_P}]$ é definido como em 32 para uma determinada distribuição P. Agora, seja P uma distribuição de probabilidade e E uma nova informação. Para mostrar que $[\mathbf{B}]_P * E = [\mathbf{B}]_{P|E}$ basta restringir a ordem de plausibilidade em $\prec_P$ aos casos onde E é o caso, e verificar que a nova ordem resultante é igual a ordem de plausibilidade em uma distribuição $P \mid E$, ou seja, é igual a ordem $\prec_{P|E}$. Uma vez que a condicionalização bayesiana sempre preserva a cotação entre as proposições, pela definição de $\prec_P$, a afirmação acima se segue automaticamente.

(iii)

Agora vamos mostrar que a teoria satisfaz Dedutivismo Hipotético e Raciocínio de Caso. Uma vez que PL rastreia condicionalização bayesiana, temos que $[\mathbf{B}]_{P|E} \models A$ *sse* $[\mathbf{B}]_P * E \models A$. Por suposição, $[\mathbf{B}]_P$ é determinado por uma ordem de plausibilidade $\prec_P$ conforme a Definição 32, o que é um caso especial do Sistema Preferencial não-monotônico. Assim, temos que, pelo axioma de *Monotonicidade Cautelosa*, o Sistema Preferencial satisfaz Dedutivismo Hipotético:

$$\text{Se } Bel(A), \text{ e } A \models E, \text{ então } Bel(A \mid E).$$

Suponha que $Bel(A)$ e $A \models E$. Uma vez que $[\mathbf{B}]$ é fechado logicamente, temos que $Bel(E)$. Pela Monotonicidade Cautelosa, temos que $[\mathbf{B}] \wedge E \models A$, e uma vez que

$[\mathbf{B}] \cap E \neq \emptyset$, temos que $[\mathbf{B}] \wedge E = Cn([\mathbf{B}] \cup \{E\}) = \mathbf{B}^+_E$, e pelo axioma ($\mathbf{B}^*4$), podemos concluir que $Bel(A \mid E)$.

De modo semelhante, a partir do axioma *Ou*, temos que *PL* satisfaz Raciocínio de Caso:

Se $Bel(A \mid E)$ e $Bel(A \mid \neg E)$, então $Bel(A)$.

Suponha que $Bel(A \mid E)$ e $Bel(A \mid \neg E)$. Pelo axioma *Ou*, temos que $E \vee \neg E \models A$. Mas sabemos que $E \vee \neg E = W$, ou seja, $Bel(A \mid W)$, que é equivalente a $Bel(A)$, satisfazendo o Raciocínio de Caso. Terminamos, assim, a demonstração das afirmações contidas no Teorema. ■

Teorema 4.3.3: Seja [**B**] o conjunto de crenças do agente, e P uma distribuição de probabilidade sobre os elementos de W. As duas afirmações a seguir são equivalentes:

(**I**) P satisfaz os postulados iniciais para probabilidade subjetiva, [**B**] satisfaz os postulados iniciais para crença, mais MR, e ambos P e [**B**] satisfazem um "Princípio de Alta Probabilidade" (PAP):

$$\text{Para todo } Y > 0, \text{ e todo } A \in \mathcal{A}, \text{ se } Bel(A \mid Y) \text{ então } P(A \mid Y) > 1/2.$$

(**II**) P satisfaz os postulados iniciais para probabilidade subjetiva. P e $\mathcal{A}$ são tais que $\mathcal{A}$ contém um menor conjunto de probabilidade 1, e existe uma classe $\mathbb{S}$ de proposições não-vazias e P–estáveis em $\mathcal{A}$ tal que: (i) $\mathbb{S}$ contém o menor conjunto de probabilidade 1 em $\mathcal{A}$, e (ii) todos outros membros de $\mathbb{S}$ possuem probabilidade menor do que 1. Além disso:

– Para todo $Y \in \mathcal{A}$ com $P(Y) > 0$: se S_1 é o menor membro de $\mathbb{S}$ no que diz respeito a relação de subconjunto, e S_1 é o menor membro de $\mathbb{S}$ para o qual temos $Y \cap S_1 \neq \emptyset$. Então, para todo $A \in \mathcal{A}$:

$$Bel(A \mid Y) \textit{ sse } A \supseteq Y \cap S_1$$

.

Demonstração. De (**I**) para (**II**):

Primeiramente, por definição P é uma distribuição de probabilidade subjetiva, tal como definimos no primeiro capítulo. Agora, para o restante da demonstração, vamos precisar de uma construção (recursiva) do conjunto $\mathbb{S}$, que será feito do seguinte modo: para todo ordinal $\alpha < \beta_P + 1$, seja

$$S_\alpha = \bigcup_{\gamma<\alpha} (S_\gamma) \cup B_{W \setminus (\bigcup_{\gamma<\alpha} S_\gamma)}$$

(onde β_P é uma boa ordem dos ordinais no que diz respeito a relação de subconjunto, de tal modo que, existe uma bijeção e um mapeamento que preserva a ordem de $\mathbb{S}_P$ em β_P. Ou seja, β_P é simplesmente uma medida do comprimento do conjunto bem-ordenado $\langle \mathbb{S}_P \rangle$).

Em particular, $S_1 = B_W$. Agora faremos algumas observações sobre a classe $\mathbb{S}$.

(a) Todo membro de $\mathbb{S}$ pertence a álgebra $\mathcal{A}$.

(b) Para todo $\gamma < \alpha < \beta_P + 1 : S_\gamma \subseteq S_\alpha$. Isso se segue diretamente da definição de membros de $\mathbb{S}$. Disso também se segue que, para dodo $\alpha + 1 < \beta_P + 1 : S_{\alpha+1} = S_\alpha \cup B_{W \setminus S_\alpha}$.

(c) Para todo $\alpha < \beta_P + 1$: $S_\alpha = \bigcup_{\gamma<\alpha} B_{W \setminus \bigcup_{\delta<\gamma} S_\delta} \cup B_{W \setminus \bigcup_{\gamma<\alpha} S_\gamma}$. Isso se segue por indução. Assuma que para todo $\gamma < \alpha$: $S_\gamma = \bigcup_{\delta<\gamma} B_{W \setminus \bigcup_{\eta<\delta} S_\eta} \cup B_{W \setminus \bigcup_{\delta<\gamma} S_\delta}$. Substitua isso pela primeira ocorrência de 'S_γ' na definição original de S_α e teremos o seguinte: $S_\alpha = \bigcup_{\gamma<\alpha}[\bigcup_{\delta<\gamma} B_{W \setminus \bigcup_{\eta<\delta} S_\eta} \cup B_{W \setminus \bigcup_{\delta<\gamma} S_\delta}] \cup B_{W \setminus (\bigcup_{\gamma<\alpha} S_\gamma)}$. Mas isso tudo pode ser simplificado pela eliminação dos ordinais contados duplamente, e teremos que: $S_\alpha = \bigcup_{\gamma<\alpha}[B_{W \setminus \bigcup_{\delta<\gamma} S_\delta}] \cup B_{W \setminus (\bigcup_{\gamma<\alpha} S_\gamma)}$, que é precisamente o que queríamos mostrar.

(d) Para todo $\alpha < \beta_P + 1$: Para todo $Y \in \mathcal{A}$, com $Y \cap S_\alpha \neq \emptyset$, temos que $B_Y \subseteq S_\alpha$. Para ver como isso é o caso, basta nota que, se $Y \cap S_\alpha \neq \emptyset$, então, pela afirmação em (c), existe um $\gamma \leq \alpha$ tal que $Y \cap B_{W \setminus (\bigcup_{\delta<\gamma} S_\delta)} \neq \emptyset$, e pela boa ordenação dos ordinais, deve existir um menor ordinal γ. Note agora que para tal menor ordinal γ temos que $Y \cap \bigcap_{\delta<\gamma} S_\delta = \emptyset$, novamente a partir de (c), e então $Y \subseteq W \setminus (\bigcup_{\delta<\gamma} S_\delta)$. Pela propriedade de Monotonicidade Racional, e uma vez que $Y \cap B_{W \setminus \bigcup_{\delta<\gamma} S_\delta} \neq \emptyset$, concluímos que $B_{[W \setminus \bigcup_{\delta<\gamma} S_\delta] \cap Y} = Y \cap B_{W \setminus \bigcup_{\delta<\gamma} S_\delta}$, o que é equivalente a $B_Y = Y \cap B_{W \setminus \bigcup_{\delta<\gamma} S_\delta}$, uma vez que $Y \subseteq W \setminus (\bigcup_{\delta<\gamma} S_\delta)$. E finalmente, dado que $Y \cap B_{W \setminus \bigcup_{\delta<\gamma} S_\delta} \subseteq B_{W \setminus \bigcup_{\delta<\gamma} S_\delta} \subseteq S_\alpha$, por (c) podemos concluir que $B_Y \subseteq S_\alpha$.

(e) Para todo $\alpha < \beta_P + 1$: S_α é *P-estável*. Isso pode ser derivado a partir da definição 22 de *P*–estabilidade: para todo $Y \in \mathcal{A}$, se $Y \cap S_\alpha \neq \emptyset$ e $P(Y) > 0$, então pela letra (d) temos que $B_Y \subseteq S_\alpha$, e então pelas propriedade de B_Y : $Bel(S_\alpha \mid Y)$. Mas isso implica que $P(S_\alpha \mid Y) > 1/2$, pela propriedade de alta probabilidade assumida em (**I**). Ou seja, S_α é *P-estável*.

(f) Existe uma menor proposição $A \in \mathcal{A}$ com probabilidade 1, $A \in \mathbb{S}$, e A é o único membro de $\mathbb{S}$ com probabilidade 1. Essa afirmação será demonstrada a partir de dois passos:

– Vamos assumir para uma redução ao absurdo que todos os conjunto S_α com $\alpha < \beta_P + 1$ possui probabilidade menor que 1. Uma vez que, por (e), todos eles são *P-estáveis*, segue-se que, a partir de (b), existe uma sequência crescente bem ordenada de conjuntos *P-estáveis* com probabilidade menor que 1, onde o comprimento de tal sequência é β_P+1. Tal sequência pode não ser *estritamente* crescente sobre os conjuntos *P-estáveis* de probabilidade menor que 1 ao longo da sequência. Pela definição de β_P, temos que β_P é propriamente o tipo ordinal de sequência de *todos* os conjuntos *P-estáveis* de probabilidade menor que 1. Então deve existir um $\alpha < \alpha' < \beta_P + 1$ tal que $S_\alpha = S_{\alpha+1}$. Então, por (b): $S_\alpha = S_\alpha \cup B_{W \setminus S_\alpha}$, e portanto, $B_{W \setminus S_\alpha} \subseteq S_\alpha$. E uma vez que $B_{W \setminus S_\alpha} \subseteq W \setminus S_\alpha$, pela definição de $B_{W \setminus S_\alpha}$, segue-se que $B_{W \setminus S_\alpha} = \emptyset$. Mas uma vez que, por suposição, $P(S_\alpha) < 1$, também segue-se que $P(W \setminus S_\alpha) > 0$, e pela condição de alta probabilidade temos que $B_{W \setminus S_\alpha} \neq \emptyset$, o que gera uma contradição. Portanto, podemos concluir que existe ao menos um conjunto S_α de probabilidade 1, sendo $\alpha < \beta_P + 1$.

– Agora na segunda parte vamos considerar um conjunto S_α qualquer, com $\alpha < \beta_P + 1$ e $P(S_\alpha) = 1$: por (c), qualquer conjunto desse tipo é a união de conjuntos da forma B_A, e qualquer um desses conjuntos S_α não pode possuir um subconjunto vazio. Mas isso significa que qualquer conjunto S_α deve ser idêntico ao único conjunto $A \in \mathcal{A}$ com probabilidade 1, o qual deve existir. Então podemos concluir que existe uma menor proposição $A \in \mathcal{A}$ com probabilidade 1, $A \in \mathbb{S}$, e A é o único membro de $\mathbb{S}$ com probabilidade 1, terminando assim a demonstração de (f).

Finalmente, a partir das afirmações (a)-(f) demonstradas acima, podemos concluir a demonstração de (**I**) para (**II**), mostrando que $Bel(A|Y)$ *sse* $A \supseteq Y \cap S_1$. Seja $Y \in \mathcal{A}$ com $P(Y) > 0$. Por (f), existe um membro de $\mathbb{S}$ com o qual Y possui uma intersecção não-vazia. Seja $\alpha < \beta_P + 1$ tal que $Y \cap S_\alpha \neq \emptyset$: por (b), S_α é o menor membro de $\mathbb{S}$ para o qual isso é o caso.

E para terminar, mostraremos que $B_Y = Y \cap S_\alpha$, o que demonstrará que a última afirmação em (**II**) se segue pela própria definição de B_Y. Considere $Y \cap S_\alpha$, que por suposição é não-vazio: por (c), $S_\alpha = \bigcup_{\gamma<\alpha} B_{W \setminus \bigcup_{\delta<\gamma} S_\delta} \cup B_{W \setminus \bigcup_{\gamma<\alpha} S_\gamma}$. Se Y possui intersecção não-vazia com qualquer conjunto da forma $B_{W \setminus \bigcup_{\delta<\gamma} S_\delta}$, com $\gamma < \alpha$, então, por (c), $Y \cap S_\gamma \neq \emptyset$. Mas isso entra em contradição com o modo como α foi definido previamente. Portanto, $Y \cap S_\alpha = Y \cap B_{W \setminus \bigcup_{\gamma<\alpha} S_\gamma} \neq \emptyset$. Em conjunto com a propriedade de Monotonicidade Racional, a última afirmação implica que $B_{[W \setminus \bigcup_{\gamma<\alpha} S_\gamma] \cap Y} = Y \cap B_{W \setminus \bigcup_{\gamma<\alpha} S_\gamma}$. E novamente pela definição de α, $Y \cap \bigcup_{\gamma<\alpha} S_\gamma$ é vazio, e assim $[W \setminus \bigcup_{\gamma<\alpha} S_\gamma] \cap Y = Y$. Portanto, temos que $B_Y = Y \cap B_{W \setminus \bigcup_{\gamma<\alpha} S_\gamma} = Y \cap S_\alpha$, chegando assim ao resultado desejado.

De (**II**) para (**I**):

Aqui a demonstração não nos exigirá tanto, e parte dela já foi apresentada ao longo do próprio texto. Primeiramente, P satisfaz os postulados iniciais para probabilidade subjetiva por hipótese. Agora, pelas propriedade de $\mathbb{S}$, sabemos que existe $S_1 \in \mathbb{S}$ tal que $S_1 = B_W$. Por hipótese, S_1 é P–estável. Pelo Teorema da Representação 3.1.1 para $TH_{\mathcal{Y}}$, apresentado no capítulo 3, temos que S_1 satisfaz os postulados iniciais para crença, uma vez que S_1 é precisamente um núcleo de crenças.

Agora precisamos mostrar que $Bel(A|Y)$ *sse* $A \supseteq Y \cap S_1$ também satisfaz (*MR*). A demonstração é análoga a do Teorema 4.2.1, já apresentada neste Apêndice. Basta notar que o lado direito da definição acima é precisamente a função de revisão de crenças em um sistema de esferas centrado em [**B**]. Agora basta notar como ($\mathbf{B}^*7$)

e (**B***8), que equivalem a propriedade MR, são satisfeitas no Teorema 4.2.1, como já apresentamos aqui.

Finalmente, vamos mostrar que a Hipótese de alta probabilidade para crença condicional apresentada em (**I**) é também satisfeita. Para isso, vamos assumir que seu antecedente é satisfeito, ou seja, $Bel(A \mid Y)$. Por (**II**) temos que $A \supseteq S_1 \cap Y$, logo, $A \cap Y \supseteq S_1 \cap Y$, o que implica, pelas propriedades da teoria da probabilidade, que $P(A \cap Y) \geq P(S_1 \cap Y)$. Como S_1 é P–estável, e $S_1 \cap Y \neq \emptyset$, temos que $P(S_1 \mid Y) > 1/2$, pela Definição 22 de P–estabilidade. Tomando esses dois fatos em conjunto, podemos afirmar que $P(A \mid Y) > 1/2$. ■