UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS HUMANAS, LETRAS E ARTES
PROGRAMA DE DOUTORADO INTEGRADO EM FILOSOFIA

UM ESTUDO LÓGICO E EPISTEMOLÓGICO DO FECHO EPISTÊMICO

Stanley Kreiter Bezerra Medeiros

João Pessoa
2013

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS HUMANAS, LETRAS E ARTES
PROGRAMA DE DOUTORADO INTEGRADO EM FILOSOFIA

UM ESTUDO LÓGICO E EPISTEMOLÓGICO DO FECHO EPISTÊMICO

Stanley Kreiter Bezerra Medeiros

Tese apresentada como requisito parcial para a obtenção do grau de Doutor em Filosofia, ao Programa Integrado de Doutorado em Filosofia, UFRN-UFPB-UFPE, CCHLA, Universidade Federal da Paraíba

Orientador: Prof. Dr. Giovanni da Silva de Queiroz

João Pessoa
2013

TERMO DE APROVAÇÃO

STANLEY KREITER BEZERRA MEDEIROS

UM ESTUDO LÓGICO E EPISTEMOLÓGICO DO FECHO EPISTÊMICO

Tese aprovada como requisito parcial para a obtenção do grau de Doutor em Filosofia, sob o título “UM ESTUDO LÓGICO E EPISTEMOLÓGICO DO FECHO EPISTÊMICO”, defendida por Stanley Kreiter Bezerra Medeiros, e aprovada em maio de 2013, em João Pessoa, Estado da Paraíba, pela Banca Examinadora constituída pelos professores:

Dr. Giovanni da Silva de Queiroz ______________________________
Orientador/UFPB

Dra. Itala Maria Loffredo D’Ottaviano ______________________________
Unicamp

Dr. José Eduardo de Almeida Moura ______________________________
UFRN

Dr. Cícero Antônio Cavalcante Barroso ______________________________
UFC

Dr. Ricardo de Sousa Silvestre ______________________________
UFCG

Dr. Anderson D’Arc Ferreira ______________________________
UFPB

João Pessoa, 23 de maio de 2013.

Agradecimentos

Agradeço ao meu orientador e amigo, Prof. Dr. Giovanni de Queiroz, pelos conselhos, pela compreensão e pelas leituras atenciosas deste trabalho (diversas, para dizer a verdade!). Agradeço especialmente a Karol, Tarski e Paco, minha família, pela companhia tão maravilhosa nos momentos de trabalho.

"Soli omnium otiosi sunt qui sapientiae vacant, soli vivunt; nec enim suam tantum aetatem bene tuentur: omne aeuum suo adiciunt." (Lucius Annaeus Seneca)

Resumo

O “fecho epistêmico” é o princípio que afirma que o conhecimento é fechado sob implicação. Se um agente S qualquer sabe que uma proposição P é o caso e, além disso, igualmente sabe que P implica logicamente outra proposição, Q, então o agente em questão também deve saber que Q é o caso. Assim, se S acredita em Q a partir da base segura fornecida pelas premissas, então ele também deve saber que Q é o caso. Este é um estudo lógico-epistemológico de princípios de fecho epistêmico com base na noção de incognoscibilidade contingente. Constatando que o problema do fecho epistêmico está em aberto e que as tentativas mais comuns na epistemologia *mainstream* contemporânea parecem ignorar os resultados da epistemologia formal sobre a relação entre estes princípios e a propriedade de onisciência lógica, nosso objetivo principal é oferecer uma estratégia para uma análise epistemológica de princípios de fecho epistêmico que considere estes resultados; que leve em conta a pretensão de aplicabilidade de um determinado princípio de fecho, segundo a situação e os agentes que se pretende modelar. Um exemplo dessa estratégia será dado ao se analisar princípios de fecho na perspectiva de agentes conjecturadores de proposições contingentemente incognoscíveis. Nossa hipótese é a de que, nesta aplicação, certos princípios de fecho não valem.

Palavras-chave: Fecho epistêmico; incognoscibilidade necessária; incognoscibilidade contingente; ceticismo.

Abstract

Epistemic closure is the principle that says that knowledge is closed under known entailment. If an agent, S, knows that some proposition P is the case and, beside of that, equally knows that P logically implies another proposition, say, Q, then S must also know that Q is the case. Thus, if S believes Q in a strong base provided by the premisses, then he must also know that Q is the case. The present work is a logical-epistemological study of epistemic closure principles, based on the notion of contingent unknowability. Noting that the problem of epistemic closure is open and that the common attempts in contemporary mainstream epistemology seem to ignore the results of the formal epistemology about the relationship between these principles and the property of logical omniscience, our main goal is to provide a strategy for an epistemological analysis of epistemic closure principles that consider these results, taking into account the purpose of applicability of a particular closure principle, according to the situation and the agents that one wants to model. An example of this strategy will be given when analyzing closure principles from the perspective of agents that reason about contingently unknowable propositions. Our hypothesis is that, in this particular application, certain closure principles do not hold.

Key-words: Epistemic closure; necessary unknowability; contingent unknowability; skepticism.

Sumário

Introdução

Desde as décadas de 60 e 70 – com as publicações de *Knowledge and Belief* (HINTIKKA, 1962) e *Epistemic Operators* (DRETSKE, 1970) – há um crescente interesse acerca do conjunto de princípios lógicos que hoje conhecemos pelo termo genérico de “fecho epistêmico”. Esse interesse se dá em pelo menos duas perspectivas: da lógica epistêmica e também da epistemologia informal. Na lógica epistêmica ou epistemologia formal, os princípios de fecho são constantemente associados a problemas sobre as capacidades cognitivas de agentes com diversos tipos de limitação – por exemplo, limitação de tempo, de recursos computacionais etc. Na epistemologia informal, esses princípios lógicos são, por sua vez, comumente associados a argumentos céticos.

O “fecho epistêmico”, grosso modo, é o princípio que afirma que o conhecimento é fechado sob implicação. Um exemplo típico e bastante discutido nas epistemologias formal e informal é o seguinte: se um agente S qualquer sabe que uma proposição P é caso e, além disso, igualmente sabe que P implica logicamente outra proposição, Q, então o agente em questão também deve saber que Q é o caso. Assim, se S acredita em Q a partir da base segura fornecida pelas premissas (ele sabe que P é o caso e que P implica Q), então ele também deve saber que Q é o caso.

Como o título sugere, este é um estudo do fecho epistêmico a partir dessas duas perspectivas diferentes (mas que são, como veremos, complementares). A primeira delas, epistemológica, procura desenvolver a problemática do fecho epistêmico segundo teorias propostas pela epistemologia *mainstream* (isto é, informal). Como iremos observar, esta perspectiva se mostra insuficiente na medida em que o conhecido “desafio de Dretske” – que lida com a questão da invalidade de alguns princípios de fecho epistêmico – pode ser considerado como não-resolvido. A constatação desse problema abre espaço para uma estratégia que vem ganhando cada vez mais força nesse campo: a consideração dos desenvolvimentos da epistemologia formal na análise de questões centrais para a epistemologia contemporânea. Uma dessas questões, e justamente a que será investigada neste trabalho, é a da relação entre certos princípios de fecho epistêmico, hipóteses céticas e proposições *heavyweight*. Apesar de sua vasta contribuição para a análise dos princípios de fecho epistêmico, a lógica epistêmica tem sido relativamente ignorada pela epistemologia *mainstream* nesse assunto. Isto nos leva à segunda perspectiva a ser considerada neste trabalho: a da lógica.

No curso desta investigação, observaremos que os resultados da epistemologia formal, representada não apenas por uma mas por diversas lógicas epistêmicas, contribuíram para uma análise cada vez mais profunda de vários princípios de fecho epistêmico. Este avanço, entretanto, se deu a partir de uma motivação inicial diferente: o estudo da relação entre princípios de fecho epistêmico e o problema da onisciência lógica. A propriedade de onisciência lógica é aquela em que um dado agente, digamos, S, é tido como conhecedor de todas as consequências lógicas daquilo que ele conhece; assim, se S conhece P e Q é uma consequência lógica de P, então S conhece Q.[1] O problema da onisciência lógica foi o responsável pelo desenvolvimento de diversas lógicas epistêmicas, especializadas exatamente em modelar situações específicas de falha de onisciência lógica, segundo suas capacidades de representação e intenções de aplicação. É justamente esta estratégia, isto é, de analisar princípios de fecho epistêmico segundo pretensões de modelagem e aplicação, uma das principais contribuições da epistemologia formal (ou lógica epistêmica) para a epistemologia *mainstream*. Tal estratégia será utilizada ao analisarmos alguns princípios de fecho na aplicação que chamamos de "conjecturadores de hipóteses céticas".

Nesse contexto, desenvolveremos a noção de incognoscibilidade necessária e incognoscibilidade contingente, e contrastaremos as duas. A diferenciação entre incognoscibilidade necessária e contingente é de extrema importância para a caracterização de hipóteses céticas e de algumas proposições *heavyweight* como "contingentemente incognoscíveis". A partir dessa caracterização, mostraremos que, na aplicação em que certos agentes epistêmicos conjecturam filosoficamente sobre hipóteses céticas, alguns princípios de fecho epistêmico são inválidos ou não-aplicáveis. Isto se dá pela relação entre "princípio de fecho epistêmico" e "proposição contingentemente incognoscível". Sustentaremos que, quando tomada como "contingentemente incognoscível", uma proposição não pode – por definição – ser conhecida, mesmo que ela seja reconhecida como uma consequência lógica de proposições que já se conhece. Este é, portanto, o resultado de nossa análise. Porém, tal como sugere a epistemologia formal, essa "invalidade" não deve ser considerada

[1] Este é um princípio de fecho conhecido em lógico epistêmica como "fecho sob implicação válida."

simpliciter, mas tão somente como sendo aplicável a essa situação específica, em que um ou mais agentes epistêmicos consideram hipóteses céticas como proposições contingentemente incognoscíveis. Isso admite a possibilidade de aplicação (ou aceitação) dos mesmos princípios de fecho em contextos ou aplicações diferentes. Assim, a própria noção informal de validade deve seguir a mesma lógica daquela da epistemologia formal: como sendo restrita a uma lógica específica ou, pelo menos, a um contexto (aplicação) específico, em que as pretensões de modelagem e aplicação estão claramente definidas.

Assim, tendo como problema central deste trabalho de investigação o fecho epistêmico e os problemas gerados por ele tanto no âmbito informal (ceticismo) quanto no formal (onisciência lógica), este é um estudo lógico-epistemológico de princípios de fecho com base na noção de incognoscibilidade contingente. Constatando que o problema do fecho epistêmico está em aberto e que as tentativas mais comuns na epistemologia *mainstream* contemporânea parecem ignorar os resultados da epistemologia formal sobre a relação entre estes princípios e a propriedade de onisciência lógica, nosso objetivo principal é oferecer uma estratégia para análise epistemológica de princípios de fecho epistêmico que considere estes resultados; isto é, que leve em conta a pretensão de aplicabilidade de um determinado princípio de fecho, segundo a situação e os agentes que se pretende modelar; um exemplo dessa estratégia será dado ao se analisar princípios de fecho na perspectiva de agentes conjecturadores de proposições contingentemente incognoscíveis (uma aplicação para a qual, como afirmamos, certos princípios de fecho não valem).

Para alcançar este objetivo, estabelecemos o cumprimento de três etapas (que serão divididas em exatamente três capítulos):

1. Primeira etapa (capítulo 1): constatação de que o conhecido “desafio de Dretske” não foi satisfatoriamente solucionado. Em outras palavras, o problema do fecho epistêmico, na perspectiva da epistemologia *mainstream*, encontra-se em aberto.

2. Segunda etapa (capítulo 2): constatação, através das soluções formais para o problema da onisciência lógica, de que o questionamento sobre validade ou invalidade de princípios de fecho, sem mais especificações sobre as pretensões de modelagem e aplicação destes, é uma simplificação exagerada da epistemologia contemporânea e impede, por sua vez, uma análise mais detalhada dos

princípios de fecho. Sugestão da utilização de recursos/estratégias da epistemologia formal na análise de princípios de fecho epistêmicos.

3. Terceira etapa (capítulo 3): caracterização de hipóteses céticas e proposições "*heavyweight*" como "proposições contingentemente incognoscíveis", e análise de princípios de fecho epistêmico segundo esta noção. Constatação da invalidade ou não-aplicabilidade de princípios de fecho epistêmico quando relacionados a proposições contingentemente incognoscíveis.

Ao término destas, esperamos ter alcançado nosso objetivo. Esperamos ter contribuído para a discussão epistemológica sobre os princípios de fecho epistêmico de pelo menos três formas: (i) um princípio de fecho não é válido ou inválido *simpliciter*, pois isso depende dos agentes e situações que se pretende modelar e, além disso, do modo que se pretende aplicá-lo; isto é, perguntar sobre a (in)validade *simpliciter* de um princípio de fecho é uma simplificação exagerada do problema. (ii) alguns princípios de fecho epistêmico não valem quando relacionados a proposições contingentemente incognoscíveis. (iii) a caracterização de hipóteses céticas e algumas proposições *heavyweight* como proposições contingentemente incognoscíveis permite ao mesmo tempo aceitar um certo nível de ceticismo para algumas proposições, mas sem cair no ceticismo radical; isso ocorre porque a incognoscibilidade de uma hipótese cética é tida apenas como contingente, e esse tipo de incognoscibilidade é bastante diferente da incognoscibilidade necessária, tal como será demonstrado no capítulo 3.

1 Fecho epistêmico: uma perspectiva epistemológica

"Não por medo, mas por dever, evitai os erros."
(Demócrates)

1.1 Introdução

Neste capítulo, o fecho epistêmico é investigado a partir de uma perspectiva epistemológica, de modo a evidenciar aquilo que, em epistemologia informal, se conhece por "o problema do fecho epistêmico". Em seguida, à luz do referido problema, investiga-se e analisa-se algumas das principais posições sobre o mesmo. Após a referida análise, termina-se por concluir que a questão do fecho ainda é um problema em aberto ou, pelo menos, que o chamado "desafio de Dretske" ainda não foi satisfatoriamente solucionado. Isto prova, portanto, que o caminho para esta tese – que procura uma nova solução para o problema do fecho, com base nas noções de "incognoscibilidade necessária" e "incognoscibilidade contingente" – ainda está aberto. Este é o principal objetivo deste capítulo.

O "fecho" – informalmente – é o princípio que afirma que o conhecimento é fechado sob implicação. Em outras palavras, se um agente S qualquer sabe que uma proposição P é caso e, além disso, igualmente sabe que P implica logicamente outra proposição Q, então o agente em questão também deve saber que Q é o caso. Assim, se S acredita em Q a partir da base segura fornecida pelas premissas (ele sabe que P é o caso e que P implica Q), então ele também deve saber que Q é o caso.

Apesar de sua plausibilidade, o princípio do fecho epistêmico tem sido submetido a inúmeras críticas e contraexemplos. Desde 1970, Fred I. Dretske é tido como um dos maiores representantes daqueles que sustentam a invalidade do fecho (para o operador de conhecimento). Segundo ele, a aceitação do referido princípio nos leva inevitavelmente ao ceticismo. Certas proposições são denominadas "céticas" quando seu conteúdo assertivo é constituído por alguma hipótese cética. Entre as proposições céticas mais famosas, podemos citar "sou uma marionete do gênio maligno" ou "sou um cérebro numa cuba". Uma hipótese cética passa a ser problemática na medida em que é considerada seriamente, ou ainda quando sua negação é a conclusão de um argumento bem construído – isto é, válido. Por exemplo: *"Sei que tenho duas mãos. Sei que, se tenho duas mãos, então não sou um cérebro numa cuba. Portanto, também sei que não sou um cérebro numa cuba"*. A forma geral do argumento acima ('sei que *P*', 'sei que *P* implica não-*Q*', 'sei que não-*Q*') é justamente uma variação de fecho epistêmico.

Observa-se – seguindo Dretske –, que o conhecimento das duas premissas do argumento acima implica no conhecimento da conclusão, a saber, "não sou um cérebro numa cuba"[1]. Porém, hipóteses céticas não são facilmente falsificáveis (pelo menos em filosofia!). Logo, a aceitação do fecho implica em ceticismo! Com base nesse raciocínio, a sugestão de Dretske é, como já foi dito, negar o fecho epistêmico, tomando-o como inválido.

Investiguemos, portanto, a problemática do fecho – não somente a posição de Dretske, mas também alguns dos argumentos em defesa do referido princípio. No entanto, antes de iniciarmos essa tarefa, devemos fornecer uma definição clara do fecho, de modo a evitar possíveis ambiguidades – tão comuns e, ao mesmo tempo, tão perigosas em assuntos filosóficos!

1.2 Definindo fecho epistêmico

Suponha que você sabe (conhece) que a seguinte proposição é verdadeira[2]:

[1]Considerando, é claro, que o fecho seja aceito.

[2]Existem, como se sabe, formas diferentes de conhecimento. Deste modo, para evitar confusões, é preciso esclarecer que interpretação damos às expressões "conhece...", "sabe...", "conhece que..." e "sabe que...". Em primeiro lugar, os verbos "conhecer" e "saber" serão utilizados, aqui, como sinônimos. Em inglês, o verbo "*to know*" é utilizado para traduzir tanto o verbo "conhecer" quanto o verbo "saber". Assim, são comuns sentenças como "*I know I'm not a brain in a vat*" ou "*Zé Eduardo knows P*" – algo que, em português, provavelmente traduziríamos por "eu sei que não sou um cérebro numa cuba" e "Zé Eduardo conhece *P*", respectivamente. Nesta última sentença, se utilizássemos "sabe" no lugar de "conhece", obteríamos "Zé Eduardo sabe *P*", uma sentença menos comum que "Zé Eduardo sabe que *P*". Colocando de outro modo: as expressões "conhece...", "sabe...", "conhece que..." e "sabe que..." serão utilizadas segundo a conveniência de tradução, na tentativa de captar a versatilidade de utilização do verbo "*to know*". O mesmo se aplica às expressões "acredita que..." e "acredita em...". Outro detalhe diz respeito às diversas formas de conhecimento. Segundo Rescher (2003, p. XIV), podemos distinguir pelo menos quatro: 1. *knowledge-that*; 2. *adverbial knowledge*; 3. *knowledge by acquaintance*; 4. *performatory knowledge*. Esta última forma refere-se

P_1: Estou lendo uma tese em Epistemologia.

Observa-se que é possível derivar um número infinito – ou pelo menos muito grande – de proposições que são consequências lógicas de P_1. Como exemplo, considere as seguintes:

P_2: Não sou um cérebro numa cuba.
P_3: Não estou lendo uma tese em Física Teórica.
P_4: Não estou jogando videogame.
P_5: Não sou uma marionete do gênio maligno cartesiano.
P_6: O mundo externo existe[3].

Não é difícil reconhecer – ao menos para pessoas racionais e em circunstâncias habituais – P_2, P_3, P_4, P_5 e P_6 como consequências lógicas de P_1. Logicamente, pode-se simplesmente dizer que estas cinco proposições são implicadas pela primeira. Formalmente, utilizando o símbolo '$\rightarrow$' para representar a implicação, escrevemos $P_1 \rightarrow P_2$, $P_1 \rightarrow P_3$ e assim por diante. Deste modo, $P_1 \rightarrow P_n$ pode ser tra-

ao conhecimento de como fazer certas coisas, ou seja, de como realizar certas tarefas. Por exemplo: Zé Eduardo sabe andar de bicicleta, dirigir, pular de paraquedas etc. O conhecimento por "*acquaintance*" é usado quando se quer destacar o conhecimento que uma pessoa tem sobre outra. Exemplo: João conhece Maria; João conhece o dono daquele celular; João sabe quem é o assassino (isto é, João conhece o assassino etc.). O conhecimento adverbial, como o nome sugere, está relacionado a advérbios. Exemplos: o conhecimento do quê, de onde, do porquê, de quando etc. A primeira forma de conhecimento que citamos, "*knowledge-that*", é aquilo que se pode chamar de conhecimento proposicional. Exemplos: Anderson sabe que está chovendo; isto é, Anderson conhece um fato sobre o atual estado de coisas. Anderson sabe que a proposição "Está chovendo." é o caso; isto é, Anderson sabe que a proposição em questão é verdadeira. Portanto, o conhecimento proposicional é o conhecimento de fatos, isto é, de que tais e tais coisas são o caso. A epistemologia lida com esta última forma de conhecimento; é a ela que nos referimos, nesta investigação, quando usamos sentenças como "S conhece P", "S acredita em Q" etc.

[3]Há duas observações importantes a fazer sobre esta sentença. A primeira é a de que, diferentemente das anteriores, ela não aparece com uma negação. A segunda é de que ela toma "existe" como predicado. Sobre a primeira observação, é importante notar que, diferentemente das demais, a proposição P_6 não é uma hipótese cética. Na seção 3.5 do capítulo 3, veremos que ela será classificada como *heavyweight*. Entretanto, similarmente às hipóteses céticas, as proposições *heavyweight* geram problemas epistemológicos quando associadas a certos princípios de fecho epistêmico. Esta é, portanto, a razão de P_6 está nesta lista. Sobre a segunda observação, a da utilização da "existência" como predicado, sabemos que uma estratégia similar foi usada por Santo Anselmo em seu famoso argumento ontológico para a prova da existência de Deus (Ver CANTUARIENSIS, Anselmus. Proslogion. The Latin Library. Disponível em http://www.thelatinlibrary.com/anselmproslogion.html.). Como sabemos desde Anselmo, sentenças como P_6, que tomam "existir" como uma "perfeição" (propriedade), trazem problemas adicionais na medida em que permitem a criação de frases do tipo "a maioria dos gregos existe, mas alguns gregos não existem"; frases como estas, como podemos notar, são destituídas de sentido. À parte deste problema, frases como "existe um mundo externo" ou "o mundo externo existe" são frequentemente usadas por epistemólogos em discussões sobre ceticismo, proposições *heavyweight* e princípios de fecho epistêmico (ver Dretske (2005a, p. 20)). Por esta razão, a proposição em questão será mantida como frequentemente aparece na literatura epistemológica.

duzida como 'P_1 implica P_n'. Do mesmo modo, utilizando 'K' como símbolo para o operador de conhecimento, escrevemos 'KP_1' para dizer 'eu conheço P_1', 'o agente S conhece P_1' ou simplesmente 'P_1 é conhecida por alguém', entre outras. Assim, pode-se formalizar uma proposição como 'eu sei que não estou lendo uma tese em Física Teórica', em nossa maneira abreviada, simplesmente escrevendo KP_3 – e assim por diante, para qualquer proposição[4]. Com isso estabelecido, a propriedade do fecho epistêmico pode ser expressa do seguinte modo:

(E-CLOS 1): Se S conhece P e conhece que $(P \rightarrow Q)$, então S conhece Q[5].

Utilizando '$\wedge$' como símbolo para a conjunção, escrevemos (E-CLOS 1) – em nosso modo abreviado – da seguinte forma: $(K_S P \wedge K_S(P \rightarrow Q)) \rightarrow K_S Q$. Hendricks (2006, p. 60) observa que o princípio do fecho epistêmico – tal como expresso acima – é equivalente ao seguinte:

(E-CLOS $1'$): Se S conhece P, então, se S conhece $(P \rightarrow Q)$, então S conhece Q[6].

Formalmente, obtemos $K_S(P \rightarrow Q) \rightarrow (K_S P \rightarrow K_S Q)$, o famoso axioma K da lógica epistêmica. Os princípios expostos acima atribuem ao operador de conhecimento a propriedade de "fecho sob implicação material"[7]. Mas que significa "fecho sob implicação"? Encontramos uma boa resposta para essa questão em Kvanvig (2006, p. 256):

> Um princípio de fecho é um princípio que afirma que uma dada categoria de objeto (tipicamente um conjunto) é fechado relativamente a alguma função, operação ou regra, de modo que a aplicação dessa operação em qualquer membro desse conjunto sempre nos levará a

[4]Quando necessário, também é possível utilizar subíndices especificadores de agentes: $K_I P_1, K_S P_1$, sendo 'I' e 'S' letras que representam o agente para quem o operador de conhecimento está sendo atribuído. De qualquer modo, é comum a omissão desses índices em casos em que a especificação é desnecessária.

[5]Optei por utilizar tal abreviação com o intuito de evitar complicações desnecessárias. Cada abreviação remete a um tipo específico de fecho epistêmico, e não outros. O termo 'E-CLOS' é uma abreviação do termo original inglês *epistemic closure*. Em seguida, associo um número natural a um tipo específico de fecho. Deste modo, a menor modificação na estrutura do fecho também resultará na modificação da sigla que o representa. A única exceção será entre os princípios (E-CLOS 1) e (E-CLOS $1'$). Para facilitar a discussão do capítulo 3, sobre a onisciência lógica, usarei os dois como sinônimos; isto é, utilizarei apenas o termo 'E-CLOS 1' para referir-me a ambos – como são princípios equivalentes, isto não irá gerar qualquer problema desnecessário.

[6]Em lógica epistêmica formal, este princípio é conhecido como "fecho sob implicação material".

[7]Na lógica epistêmica clássica, 'implicação lógica' e 'implicação material' coincidem. Nas deliberações epistemológicas deste capítulo, utilizarei frequentemente as expressões "*fecho sob implicação*" e "*fecho epistêmico*" para designar o princípio geral de que, "se o agente S conhece P e conhece $P \rightarrow Q$, então o agente S conhece Q".

> algo que já se encontra no conjunto. Considere, por exemplo, os números naturais e a operação de adição. Metaforicamente falando, podemos descrever a porta desse conjunto como fechada sob essa operação, dado que, sempre que adicionamos dois números naturais, isso nos leva a algo que já se encontra no conjunto[8].

No caso do fecho epistêmico – ou fecho do operador de conhecimento – a propriedade definida por Kvanvig é supostamente satisfeita. Para compreender isso, considere K não como um operador modal de conhecimento, mas como uma classe de conhecimento. Ou seja, entenda K como "a classe das coisas que são conhecidas". Seja S um agente e P uma proposição qualquer tal que S conhece P. Assim, afirmar "S conhece P" é simplesmente afirmar que P pertence a K_S – isto é, à classe das proposições conhecidas por S. Suponha também que $(P \to Q) \in K_S$ – isto é, que S conhece $(P \to Q)$. Agora, para expressar o fecho epistêmico, escrevemos:

- Se $P \in K_S$ e $(P \to Q) \in K_S$, então $Q \in K_S$.

Observa-se, portanto, que afirmar o fecho epistêmico é simplesmente afirmar que a função implicação, quando aplicada aos membros da classe das coisas conhecidas, gera elementos que já pertencem à referida classe. Princípios como (E-CLOS 1) e (E-CLOS 1') são, na verdade, apenas exemplos particulares de fecho epistêmico. Alguns são mais fortes, outros mais fracos; há também aqueles que seriam supostamente tipos de fecho epistêmico, quando na verdade não o são. Os princípios acima, bastante discutidos em lógica epistêmica e epistemologia, também são os mais problemáticos em ambas as áreas. A eles associamos, em lógica epistêmica, o conhecido problema da "onisciência lógica" (HINTIKKA, 1962). Em epistemologia, o problema da aceitação do ceticismo (DRETSKE, 1970) é um deles.

1.3 Investigando o fecho epistêmico

Considere a seguinte instância de (E-CLOS 1):

[8]"*A closure principle is a principle that claims that a certain category of object (typically a set) is closed relative to some function or operation or rule, in the sense that performing that operation on any member of the set always leads us to something already in the set. Consider, for example, the natural numbers and the operation of addition. Speaking metaphorically, we might describe the door to this set as being closed under this operation, since whenever we add two natural numbers, we are led to something already in the set.*"

Instância 1: $(K_V P_1 \wedge K_V(P_1 \rightarrow P_4)) \rightarrow K_V P_4$[9].

Ou, informalmente:

- $K_V P_1$: Você sabe que você está lendo uma tese em epistemologia.
- $K_V(P_1 \rightarrow P_4)$: Você também sabe que, se está lendo uma tese em epistemologia, então você não está jogando videogame.
- $K_V P_4$: Portanto [pelo fecho epistêmico], você sabe que não está jogando videogame.

Ao observar cuidadosamente o argumento acima exposto, não é de todo estranho que associemos a ele as seguintes questões[10]:

1. Há alguma coisa errada com este princípio?
2. Devo realmente saber todas as consequências lógicas daquilo que conheço? Isto é, o meu conhecimento é realmente fechado sob implicação?
3. Como questionou Dretske (1970), é de fato o operador de conhecimento "plenamente penetrante"[11]?

Antes de nos debruçarmos sobre essas questões, porém, discorramos um pouco mais sobre os tipos de fecho e suas respectivas áreas de interesse.

Como dissemos, além de suas versões epistemológicas – bem como dos problemas associados às mesmas em epistemologia – os princípios de fecho epistêmico são geralmente associados com a lógica epistêmica e com o conhecido problema da onisciência lógica. Na verdade, a utilização de termos como "princípio de onisciência lógica" para designar um certo tipo de fecho do operador de conhecimento

[9]No presente caso, o 'V' entre 'K' e as proposições 'P_1', '$(P_1 \rightarrow P_4)$' e 'P_4' serve como marcador para o agente a quem o operador de conhecimento está sendo atribuído. Assim, 'V' representa, neste caso particular, você mesmo (isto é, o leitor desta tese em epistemologia). Deste modo, quando afirmo '$K_V P_1$', estou atribuindo o conhecimento da proposição 'P_1' a você.

[10]Tal atitude é justamente a mesma tomada por Dretske (1970) e Almeida (2007). Ambos negam o fecho epistêmico, diferenciando-se entre si no que se refere aos argumentos adotados em sua invalidação. No entanto, analisando a posição dos referidos autores, uma coisa pode ser afirmada: questionar o fecho epistêmico não é algo necessariamente absurdo. A atitude de negá-lo pode, ao contrário, ser algo bastante natural – como veremos no decorrer deste capítulo e, de um modo geral, desta tese.

[11]O termo original encontrado no texto "*Epistemic Operators*" (DRETSKE, 1970) é "*fully penetrating operator*".

é bastante comum em lógica epistêmica. Quando inseridos nesse contexto, esses princípios são, por sua vez, de grande interesse para lógicos e pesquisadores em IA (inteligência artificial). De fato, todas as formas de fecho listadas abaixo são bastante importantes para esses teóricos[12]:

(E-CLOS 2) [onisciência lógica total]: Um agente é total-logicamente onisciente se, sempre que ele conhece todas as fórmulas de um conjunto Γ *e* Γ *implica logicamente a fórmula* φ*, então o agente também conhece* φ*.*

(E-CLOS 3) [conhecimento de fórmulas válidas]: Se φ *é válida, então o agente* S *conhece* φ*. Isto é, se* $\models \varphi$*, então* $K_S\varphi$*.*

(E-CLOS 4) [fecho sob implicação lógica]: Se o agente S *conhece* φ *e* φ *implica logicamente* ψ*, então o agente* S *conhece* ψ *. Isto é, se* $K_S\varphi$ *e* $\varphi \models \psi$*, então* $K_S\psi$*.*

(E-CLOS 5) [conhecimento sob equivalência lógica]: Se o agente conhece φ *e se* φ *e* ψ *são logicamente equivalentes, então o agente* S *conhece* ψ*. Ou seja, se* $K_S\varphi$ *e* $\models \varphi \leftrightarrow \psi$*, então* $K_S\psi$*.*

(E-CLOS 6) [fecho sob implicação válida]: Se o gente S *conhece* φ *e* $(\varphi \rightarrow \psi)$ *é valida, então o agente* S *conhece* ψ*. Isto é, se* $K_S\varphi$ *e* $\models (\varphi \rightarrow \psi)$*, então* $K_S\psi$*.*

(E-CLOS 7) [fecho sob conjunção]: Se o agente S *conhece ambas* φ *e* ψ*, então o agente* S *conhece* $(\varphi \wedge \psi)$*. Isto é, se* $K_S\varphi$ *e* $K_S\psi$*, então* $K_S(\varphi \wedge \psi)$*.*

(E-CLOS 8) [fecho sob distribuição da conjunção]: Se o agente S *conhece a conjunção de* φ *e* ψ*, isto é,* $(\varphi \wedge \psi)$*, então* S *conhece* φ *e conhece* ψ*. Ou seja, se* $K_S(\varphi \wedge \psi)$*, então* $K_S\varphi$ *e* $K_S\psi$[13]*.*

Note-se que (E-CLOS 2) é um princípio muito forte. A intuição por trás dele é a de que o agente S conhece todas as consequências lógicas daquilo que ele conhece. Assim, se S conhece todas as proposições do conjunto G, então qualquer

[12]Para um estudo detalhado sobre onisciência lógica, consultar Fagin *et al* (2003): *Reasoning about knowledge*. Massachusetts: MIT Press, capítulo 9. Os princípios apresentados são expressos em sua forma lógica original (p. 335), daí a utilização de letras gregas. Eles são enunciados em metalinguagem, de modo que as referidas letras atuam como marcadores para as fórmulas da linguagem objeto.

[13]Escolhi adicionar este princípio à lista porque ele será de grande importância para nossas deliberações. De fato, um dos argumentos de Hawthorne (2005, p. 31-32) contra a posição de Dretske (2005a e 2005b) é o de que, ao adotarmos a posição deste último com relação ao fecho epistêmico, estaremos imediatamente comprometidos com a rejeição do princípio da distribuição da conjunção. Tal atitude não seria sábia, segundo Hawthorne, dado que "a distribuição [da conjunção] parece incrivelmente plausível".

fórmula logicamente implicada por G também é conhecida por S. No caso $G = \emptyset$, dizemos então que S conhece todos os teoremas. Em outras palavras, o agente S é total-logicamente onisciente. Apesar de seus vários usos e aplicações, o princípio (E-CLOS 2) provavelmente não será aceito caso o objetivo dos lógicos ou dos pesquisadores em inteligência artificial seja modelar o raciocínio limitado e, portanto, as capacidades cognitivas reais de agentes com recursos cognitivos limitados. É justamente por isso que, em lógica epistêmica, há tantas abordagens aptas a invalidarem o referido princípio[14].

Epistemologicamente falando, a onisciência lógica total não é uma propriedade plausível. Como sabemos, os seres humanos são cognitivamente limitados de várias maneiras. Não somos logicamente oniscientes. Em algumas circunstâncias, não somos capazes de derivar consequências lógicas do nosso próprio conhecimento. Uma observação importante sobre (E-CLOS 2) é a de que, segundo tal princípio, um agente S conhece ψ mesmo que ele não esteja ciente da implicação relevante de φ para ψ. Essa não é uma forma realista de ver as capacidades reais de agentes cognitivamente limitados, dado que, em muitas ocasiões, falhamos em conhecer uma proposição justamente porque não estamos cientes de que ela é uma consequência lógica daquilo que sabemos; isto é, justamente porque não temos ciência da implicação relevante. Por essa razão, uma teoria epistemológica menos ambiciosa está comprometida com a rejeição de (E-CLOS 2). Outras razões em favor da rejeição de (E-CLOS 2), para agentes cognitivamente limitados, podem incluir:

(1) **Recursos limitados** (os agentes são *resource-bounded*). Alguns agentes possuem recursos computacionais limitados para derivar consequências lógicas daquilo que eles já sabem.

Para deduzir ψ de sua base de conhecimento (daquilo que ele sabe), um agente S pode necessitar de alguma regra de inferência específica R de tal modo que, se R não estiver à disposição de S, S não será capaz de inferir ψ. Em outras palavras, a inferência em questão não poderá ser efetuada porque alguma regra de inferência está faltando; não há outra alternativa para S. Chamemos essa categoria de regras de inferência de regras-chave.

[14] Desde o trabalho pioneiro de Hintikka em lógica epistêmica – *Knowledge and Belief* (1962) – muitas "soluções lógicas" foram propostas para o problema da onisciência lógica. Já em "*Knowledge and Belief*" Hintikka está ciente do problema. Para algumas das chamadas "soluções lógicas", consultar Hintikka (1975), Fagin e Halpern (1988) e também Fagin et al (1995). Há também outros trabalhos relevantes sobre o tema em Levesque (1984) e Konolige (1984, 1986a e 1986b).

Definição 1.1. *[regra-chave]: Seja S um agente, R uma regra de inferência e* $G_S \vdash \psi$ *uma dedução de* ψ *a partir da base de conhecimento G do agente S. R é uma regra chave para um agente S se, e somente se, R é indispensável para S derivar* ψ.

Há também outras formas de limitação de recursos. Ainda que S possua à sua disposição todas as regras de inferência de que ele necessita para deduzir uma proposição ψ a partir do que ele sabe, pode ser o caso que, em uma circunstância específica, ele não tenha à disposição tempo suficiente para fazê-lo. Como sabemos, o ato de realizar deduções é parte de um processo de inferência. Ora, é bastante razoável supor que os processos de inferência necessitam de uma certa quantidade de tempo para se completarem. Logo, em muitos casos, para que um agente S possa deduzir uma dada proposição de sua base de conhecimento, ele precisará de algum tempo mínimo – tempo que seja, ao mesmo tempo, necessário e suficiente para a realização da dedução em questão.

Com base nisso, considere um processo de inferência em que o tempo disponível para se deduzir uma proposição verdadeira ψ, de uma dada base de conhecimento G, é t. Imagine agora uma circunstância em que um dado agente, S, tem de deduzir uma proposição ψ de sua base de conhecimento. Considere que, para realizar a dedução em questão, S necessita do tempo $t+1$. Inevitavelmente, S não será capaz de deduzir ψ a partir de G devido aos recursos limitados de tempo que estão à sua disposição. O caso pode ser ilustrado do seguinte modo: um aluno de matemática em dia de prova.

Suponha que um aluno do ensino médio, João, esteja prestes a fazer uma prova de matemática. O professor lhe entrega a prova e o lembra da duração máxima da atividade avaliativa: 2:00 horas. João, apesar de muito inteligente, é também bastante preguiçoso. Entre as questões da prova, há uma que tem como resposta uma dada proposição ψ. Suponha que, para derivar ψ, João precisaria – devido a sua preguiça e, consequentemente, seu ritmo lento – de pelo menos 2:30 horas. Consequentemente, no caso em questão, João possui recursos limitados de tempo para a derivação de ψ, de modo que não será capaz de efetuar tal derivação em tempo hábil. Repare que, mesmo considerando que João saiba implicitamente a resposta para essa questão – dado que possui um conjunto de proposições e regras que tem ψ como consequência lógica – não nos é permitido afirmar que João conhece a proposição ψ – ao menos não explicitamente. Logo, neste caso particular – em que o tempo para a derivação não está disponível –, o agente em questão falha na dedução de ψ, mesmo que ψ seja uma consequência lógica de sua base de conhecimento.

(2) **Ausência de consciência de conceitos relevantes**. Certamente, você não pode afirmar que conhece ou desconhece uma proposição qualquer P, se P requer – de sua parte – consciência (ou ciência) de um conceito que lhe é completamente alheio.

Isso ocorre mesmo se P for uma consequência lógica daquilo que você conhece. Há um bom exemplo para ilustrar esse caso. Imagine que alguém vem até você e pergunta: "Você sabe se aquele cantor, Sting, também é um filósofo?" A resposta para essa questão talvez seja bem fácil; basta apenas conhecer – mesmo que vagamente – o conceito de "filosofia". Contudo, se você não apreendeu tal conceito, ou se ele estiver completamente indisponível para você na ocasião em questão, então você estará fortemente inclinado a dizer que não é capaz de responder a referida pergunta, dado que não tem a menor ideia do que seja "filosofia". Logo, em casos como esse – e também considerando que você é uma pessoa cognitivamente coerente –, você não poderá dar nem uma resposta positiva nem uma resposta negativa à pergunta. O melhor seria, simplesmente, suspender o juízo.

Ao investigar o fenômeno da incognoscibilidade, Rescher (2009, p. 8) escreve algo que, além de muito interessante, está bem próximo desse caso particular de falha em onisciência lógica:

> Júlio César não poderia ter se perguntado se sua espada continha tungstênio, ou se Rutherford B. Hayes venceu a eleição presidencial dos EUA legalmente. Os próprios conceitos necessários para formular tais questões estão fora dos horizontes conceituais de pessoas desses tempos e lugares – ou, possivelmente, de pessoas de todos os tempos e lugares[15].

Observa-se, com isso, que algumas questões são simplesmente irrespondíveis. Isso ocorre porque, como Rescher observa, os conceitos necessários para compreendê-las não estão ao nosso alcance. Da mesma forma, algumas proposições nos são desconhecidas justamente porque não estamos cientes dos conceitos relevantes necessários à sua compreensão.

[15]"*Julius Caesar could not have wondered if his sword contained tungsten or if Rutherford B. Hayes won the U.S. presidency legally. The very concepts needed to form such question are outside the conceptual horizons of people at some times and places – or possibly of people at all times and places.*"

(3) **Preconceito**. Um agente qualquer pode falhar em conhecer uma proposição *P* (ou até mesmo acreditar em *P*) devido aos preconceitos que alimenta. Novamente, pode acontecer de *P* ser uma consequência lógica daquilo que ele sabe. Ele pode até estar consciente de que *P* segue daquilo que ele conhece, mas não aceita o fato devido aos seus próprios preconceitos.

Para ilustrar o caso, suponha que um neonazista – depois de alguns estudos e observações –, deduz a seguinte proposição de sua (nova) base de conhecimento: "Minhas ideias neonazistas sobre a humanidade estão completamente equivocadas". Considerando o fato de essa proposição não poder ser aceita pelo nosso agente devido a suas crenças pessoais, muitos defenderiam a tese de que ele não conhece a proposição em questão. Se, por acaso, defendermos a tese de que ele a conhece, então estaremos comprometidos com a aceitação de:

- *O agente S conhece P, mas não aceita P.*

Que seria muito similar a:

- *O agente S conhece P, mas não acredita que P.*

Esta última, por sua vez – quando formulada em primeira pessoa –, nos dará:

- *M: P, mas não acredito (não aceito) que P.*

Consequentemente, estaremos comprometidos com a aceitação de uma proposição de Moore, caracterizada por ser paradoxal[16]. Em casos como este, no entanto, alguns diriam que os psicólogos teriam respostas mais interessantes que os epistemólogos[17]!

Até agora, vimos várias razões contra a aceitação de (E-CLOS 2). Obviamente, não apresentamos todas aqui; há muitas outras, lógicas e epistemológicas. Mas que dizer dos outros princípios de fecho epistêmico? Ora, os princípios (E-CLOS 3), (E-CLOS 4) e (E-CLOS 6) são muito próximos a (E-CLOS 2) – apesar de

[16]Para uma excelente análise do paradoxo de Moore e mais um caso contra o fecho epistêmico, ver DE ALMEIDA. Racionalidade epistêmica e o paradoxo de Moore. **Veritas.** Vol. 54, n. 2, p. 48-73, maio/ago. 2009.

[17]Um exemplo similar (apesar de não equivalente) envolve a noção de "ansiedade epistêmica"; tal exemplo é fornecido por Hawthorne (2005, p. 40).

não possuírem a mesma força. No entanto, observamos que, apesar da semelhança, eles não são necessariamente dependentes um do outro. Em Fagin *et al* (2003), encontramos a ideia de que a aceitação/rejeição de cada um depende da aplicação da lógica utilizada. Em outras palavras, não é obrigatório para um lógico que rejeita (E-CLOS 2) também rejeitar os demais. Sistemas lógicos construídos para invalidar (E-CLOS 2) podem, ao mesmo tempo, validarem (E-CLOS 3), (E-CLOS 4) e (E-CLOS 5) (entre outros). Novamente: tudo irá depender da aplicação da lógica a ser utilizada[18].

Do simples ponto de vista de uma epistemologia descritiva, (E-CLOS 3), (E-CLOS 4) e (E-CLOS 6) também são inaceitáveis. Eles são muito generosos com os seus agentes. Agentes reais nem sempre possuem as capacidades lógicas que esses princípios lhes atribuem. Para isso, analisemos uma versão epistemológica aproximada de (E-CLOS 3):

(M-CLOS 3) [conhecimento de "verdades"]: Se t é uma verdade, então o agente S conhece t[19].

A intuição por trás deste princípio é a de que o agente em questão conhece todas a verdades lógicas. Certamente, essa tese é incompatível com a ideia de um agente cognitivamente limitado. Observações similares se aplicam às versões epistemológicas de (E-CLOS 4) e (E-CLOS 6). O princípio (M-CLOS 3) é, além disso,

[18] A ideia de endossar/rejeitar algum princípio particular de fecho em virtude da aplicação da lógica de interesse é algo normal e bem conhecido pelos lógicos. Naturalmente, uma lógica que tenta modelar as capacidades computacionais de um robô não precisa necessariamente validar os mesmos princípios de fecho de outra que se interessa por seres humanos, ou outra que investiga agentes com poderes computacionais "extra-humanos". Isso é justamente o que tentaremos estabelecer no segundo capítulo, juntamente com a ideia de que o mesmo procedimento – a saber, a análise dos princípios de fecho em virtude de sua aplicação – deveria ser adotado pela epistemologia *mainstream*. De fato, sustentamos que a discussão epistemológica sobre o fecho só chegou a essa confusão em que está hoje em dia porque seus contendores tendem a simplificar demais o problema do fecho, estabelecendo dicotomias obrigatórias como "validade ou invalidade" do fecho – sendo este "ou" forçada e erroneamente exclusivo. Nossa perspectiva é de que essa dicotomia é facilmente superada com a adotação da metodologia já existente na própria lógica epistêmica, que é de pensar que a validade ou invalidade dos variados princípios de fecho dependem da aplicação da lógica de base.

[19] Primeiramente , (M-CLOS) pode ser tomado como uma abreviação para o termo *Mainstream Closure*, de *Mainstream Epistemology*. Utilizarei esta abreviação para princípios de fecho em formato epistemológico (informal), diferentemente dos (E-CLOS), em cujos enunciados encontramos referências a fórmulas e outros termos típicos da lógica formal. Deste modo, pense em (M-CLOS) como uma versão epistemológica para (E-CLOS). Além disso, observe-se também que t não é uma simples proposição contingentemente verdadeira. Se isso fosse o caso, (M-CLOS 3) poderia ser convertido no axioma lógico $P \to K_S P$, inaceitável até mesmo na lógica epistêmica clássica. Pense em t simplesmente como uma constante para uma verdade lógica (uma verdade invariável). Poderíamos utilizar o termo "verdade necessária" para t. Contudo, para evitar maiores complicações filosóficas, evitemos esse caminho, se possível.

incompatível com a ideia de que existem "verdades incognoscíveis" e "implicações incognoscíveis" ("verdades" e "verdades logicamente implicadas" que ninguém pode possivelmente conhecer)[20].

O que dizer das versões epistemológicas de (E-CLOS 7) e (E-CLOS 8)? Isto é, o que dizer de:

(M-CLOS 7) Fecho sob conjunção: Se o agente S conhece P e conhece Q, então o agente S também conhece $(P \wedge Q)$.

(M-CLOS 8) [Fecho sob distribuição a conjunção]: Se o agente conhece a conjunção $(P \wedge Q)$, *então o agente S conhece P e também conhece Q.*

Neste exato momento, parece bem razoável para você, leitor, admitir que você conhece ao menos duas proposições:

1. *Eu estou vivo.*
2. *Estou lendo uma tese em epistemologia.*

Ora, também parece bastante razoável que, se você conhece '1' e também conhece '2', então você conhece a conjunção dessas duas proposições, a saber:

3 *Estou vivo e estou lendo uma tese em epistemologia.*

Apesar de sua plausibilidade, o princípio (M-CLOS 7) não está isento de críticas. Uma abordagem probabilística sobre o conhecimento, juntamente com (M-

[20] A existência de proposições necessariamente (ou logicamente) incognoscíveis é demonstrada em Fitch (1963). Há também um resultado de 1984 no qual MacIntosh demonstra a incompatibilidade entre as três teses abaixo:

1. Distribuição ou fecho da conjunção: Se $K_S\varphi$ e $K_S\psi$, então $K_S(\varphi \wedge \psi)$,
2. Veracidade: Se S conhece P, então P é verdadeira P $(K_SP \rightarrow P)$.
3. Cognoscibilidade: Se P é verdadeiro, então é possível que o agente S conheça P.

O resultado é conhecido na literatura como "teorema do colapso", que por sua vez estabelece que: se juntarmos (3), (1) e (2), então teremos o resultado de que "S conhece P se, e somente se, P for verdadeira" $(K_SP \leftrightarrow P)$. Assim, para qualquer agente S e qualquer proposição P, se P for verdadeira, então S conhece P. Mas sabemos que este último enunciado é incompatível com o teorema de Fitch de que existem proposições que são, ao mesmo tempo, verdadeiras e incognoscíveis. Mais tarde (capítulo 3), utilizarei as noções de "incognoscibilidade necessária" e "incognoscibilidade contingente" para mostrar uma nova forma de invalidação do fecho epistêmico.

CLOS 7), gera um inconveniente conhecido na literatura epistemológica como "o paradoxo da loteria", a ser exemplificado a seguir[21].

Suponha que você apoia uma abordagem probabilística qualquer para a noção de conhecimento[22]. Agora, suponha ainda que a condição para que você conheça uma dada proposição *P* é a que, além de acreditar em *P* e asseri-la, a probabilidade de *P* ser verdadeira precisa ser muito próxima a 1 (apesar de não precisar ser exatamente 1). Adotando essa perspectiva, ser-lhe-á permitido afirmar que o bilhete de loteria que você acabou de comprar não será o premiado. Na verdade, a probabilidade de você perder a loteria é tão alta que, sem muitos problemas, seria até permitido afirmar que você sabe que seu bilhete não será o premiado. O problema, portanto, é gerado: de fato, você pode afirmar, de cada bilhete em particular, que ele não será o premiado – isto é, considerando cada bilhete separadamente. Porém, pelo princípio (M-CLOS 7), se você sabe – separadamente – que cada bilhete não será o premiado, então você também conhece a conjunção desses fatos, a saber, que nenhum bilhete será premiado! Mas é óbvio que isso contradiz o fato de que haverá ao menos um bilhete premiado. O resultado, como você pode ver, é um paradoxo: "o paradoxo da loteria".

Uma possível solução para esse paradoxo seria simplesmente rejeitar que alguém possa acreditar justificadamente, de um bilhete particular de loteria, que tal bilhete não será o premiado. Também é possível desenvolver diferentes teorias bem sucedidas com o propósito de eliminação de paradoxos de loteria[23].

Agora, diferentemente de (M-CLOS 7), o princípio (M-CLOS 8) já não é assim tão problemático. Se eu conheço a conjunção '*estou vivo e lendo sobre o fecho epistêmico*', então não há problema em afirmar que eu sei esses dois fatos separadamente: (1) '*eu estou vivo*' e (2) '*eu estou lendo sobre o fecho epistêmico*'. O princípio (M-CLOS 8) é tão intuitivo que até mesmo aqueles que rejeitam certos princípios de fecho epistêmico – como é o caso de Fred Dretske – parecem aceitá-lo. Em seu artigo monumental de 1970, *Epistemic operators* (p. 1009), Dretske escreve:

> [...] parece bastante óbvio que, se alguém conhece *P* e *Q*, tem razão para acreditar que *P* e *Q*, ou pode provar que *P* e *Q*, esse alguém

[21]Para as versões originais do referido paradoxo, ver KYBURG, Henry L. **Probability and the logic of rational belief**. Middletown: Wesleyan University Press, 1961; Conjunctivitis. *In*: Swain, M. (editor), **Induction, acceptance and rational belief**. Dordrecht: Reidel, 1970, p. 55-82.

[22]Pense, por exemplo, que cada proposição possui um nível de probabilidade que vai de 0 (impossível) a 1 (certa).

[23]Ver DOUVEN, Igor. The lottery paradox and our epistemic goal. **Pacific Philosophical Quarterly**. Vol. 89, n. 2, p. 204-225, jun. 2008.

> consequentemente conhece que P, tem razão para acreditar que Q, ou pode provar (no sentido apropriado deste termo) que Q. Similarmente, se S sabe que Bill e Susan casaram um com o outro, ele (deve) saber que Susan se casou (casou com alguém)[24].

Contudo, como mostra Hawthorne (2005, p. 31), a abordagem de Dretske acerca do conhecimento "...está comprometida com a rejeição da distribuição". Examinarei as posições de Dretske e Hawthorne nas próximas seções. Por hora, voltemos nossa atenção para o princípio (E-CLOS 1)[25].

Muitos diriam que este princípio é, sem dúvida, um dos mais plausíveis dentre aqueles que discutimos até agora. Lembre-se do exemplo fornecido na instância 1 (seção 1.3). Neste exato momento, você está lendo uma tese em epistemologia. Se você é do tipo de pessoa que não consegue fazer várias coisas ao mesmo tempo, então você certamente não está jogando *videogame* – ao menos enquanto estiver lendo esta tese em epistemologia.

Neste exato momento, você consegue ver (perceber) a si mesmo lendo esta tese em epistemologia. Você pode ver o texto à sua frente. Depois de alguns momentos de reflexão, você se dá conta de que, se você está lendo esta tese em epistemologia, então você não está jogando *videogame*. Logo, lhe parece óbvio, com base nessas informações, que você não está jogando *videogame*. De fato, você não está jogando *videogame*! Que princípio epistemológico maravilhoso! Poderia um princípio tão simples e plausível estar errado?

Como veremos em breve, o princípio de fecho epistêmico (E-CLOS 1), como uma tese, não é tão forte como o infalibilismo, mas é suficiente para apoiar a conclusão cética de que não temos conhecimento nem mesmo de hipóteses (proposições) comuns como '*eu estou lendo, neste momento, uma tese em epistemologia*'. Deste modo, para provar que o conhecimento é realmente possível, tanto Dretske (1970) quanto Nozick (1981, p.204-211) pensaram que deveríamos rejeitá-lo, e ambos nos forneceram razões interessantes para fazê-lo. Dretske (2005a, p.18) vê a rejeição do fecho epistêmico como a única forma de escapar ao ceticismo, dado que "[...] o único meio de preservar o conhecimento de verdades caseiras, as verdades que

[24]"[...] *it seems to me very obvious that if someone knows P and Q, has a reason to believe that P and Q, or can prove that P and Q, he thereby knows that Q, has a reason to believe that Q, or can prove (in the appropriate sense of the this term) that Q. Similarly, if S knows that Bill and Susan married each other, he (must) know that Susan got married (married someone).*"

[25]Discutirei o princípio (E-CLOS 5) mais à frente, quando estiver lidando com os argumentos de Hawthorne contra a teoria das razões conclusivas de Dretske.

todos julgam conhecer, é [...] abandonar o fecho”[26].

1.4 O argumento cético por *Modus Tollens*

Considere novamente as proposições da seção anterior:

P_1: Estou lendo uma tese em Epistemologia.

P_2: Não sou um cérebro numa cuba.

P_3: Não estou lendo uma tese em Física Teórica.

P_4: Não estou jogando videogame.

P_5: Não sou uma marionete do gênio maligno cartesiano.

P_6: O mundo externo existe.

Agora, suponha que a proposição P_1 seja o caso; isto é, suponha que você esteja, neste momento, lendo uma tese em epistemologia. Dado isto, qualquer pessoa razoável diria que P_1 implica P_2, bem como P_5. Assim, se estou lendo uma tese em epistemologia, então eu não sou um cérebro numa cuba, em um planeta chamado “*alfa*”! Similarmente, se estou lendo uma tese em epistemologia neste momento, então certamente não estou sendo enganado pelo gênio maligno cartesiano. Deste modo, não há nada estranho em aceitar P_2 e P_5 como consequências lógicas de P_1. Lembre-se do fecho epistêmico, e considere agora estas duas instâncias:

Instância 2: $(K_V P_1 \wedge K_V(P_1 \rightarrow P_2)) \rightarrow K_V P_2$

Informalmente:

Você sabe que está lendo uma tese em epistemologia.

Você também sabe que, se está lendo uma tese em epistemologia, então não é um cérebro numa cuba.

Logo – pelo fecho epistêmico – você sabe que não é um cérebro numa cuba.

Instância 3: $(K_V P_1 \wedge K_V(P_1 \rightarrow P_5)) \rightarrow K_V P_5$

Informalmente:

Você sabe que você está lendo uma tese em epistemologia.

[26]O termo original, em inglês, é *homely truths*: “[...] The only way to preserve knowledge of homely truths, the truths everyone takes themselves to know, is [...] to abandon closure.”

Você também sabe que, se está lendo uma tese em epistemologia, então não é uma marionete do gênio maligno cartesiano.

Logo – pelo fecho epistêmico – você sabe que não é uma marionete do gênio maligno cartesiano.

Proposições como P_2 e P_5 são conhecidas na literatura epistemológica como "hipóteses céticas" ou "possibilidades de erro céticas"[27]. Chamamo-las assim justamente porque – ao menos à primeira vista – elas representam possibilidades que bloqueiam qualquer pretensão de conhecimento. Assim, quando você (verdadeiramente) diz "*estou, neste exato momento, lendo uma tese em epistemologia*", o cético pode atacar sua afirmação com "*você está certo disso, considerando a possibilidade de que, ao invés de estar lendo uma tese em epistemologia, você seja uma marionete do gênio maligno cartesiano?*". Deste modo, parece que, se as proposições "*não sou um cérebro numa cuba*" e "*não sou uma marionete do gênio maligno cartesiano*" são consequências lógicas de "*estou, neste exato momento, lendo uma tese em epistemologia*", então você também deveria saber que:

1. Você não é um cérebro numa cuba.

2. Você não é uma marionete do gênio maligno cartesiano.

Assim, supondo que você conhece (i) que você está lendo uma tese em epistemologia e que também conhece (ii) as implicações $(P_1 \rightarrow P_2)$ e $(P_1 \rightarrow P_5)$, então você também deveria conhecer, ambas, P_2 e P_5. Como você já deve ter observado, isso é exatamente o que é requerido pelo fecho epistêmico. Assumindo que seja verdade que você sabe que está lendo uma tese em epistemologia e que, '*estar lendo uma tese em epistemologia*' implica '*não ser um cérebro numa cuba*', então você deve saber que não é um cérebro numa cuba (o mesmo vale para o caso do gênio maligno). O cético pode agora fazer sua jogada.

Ele pode gentilmente lhe pedir para considerar a possibilidade de ser um cérebro em uma cuba – ou uma marionete do gênio maligno cartesiano. No atual estado de coisas, não há como você saber que você não é uma cérebro numa cuba – diria o cético. Se você fosse um cérebro numa cuba, você não teria como saber de sua

[27]Na verdade, nas instâncias 2 e 3 apresentadas até agora, P_2 e P_5 estão representando as negações das hipóteses céticas "*sou um cérebro numa cuba*" e "*sou uma marionete do gênio maligno cartesiano*". Decidi não utilizar o símbolo lógico da negação, até agora, com o intuito de facilitar a compreensão dos exemplos. O símbolo '$\neg$' será introduzido em momento oportuno.

real situação – ao menos não sozinho, isto é, sem a ajuda de terceiros. Cérebros em cubas são estimulados por computadores poderosos a terem “as mesmas” sensações e percepções que as pessoas normais têm – simulando uma espécie de realidade virtual[28]. Deste modo, como dito antes, não há como você saber que não é um cérebro numa cuba.

Esta é uma premissa importante para o argumento cético. Ele, o cético, deseja que você a aceite como uma genuína possibilidade de erro. Assim, para ter conhecimento de qualquer coisa – diria ele – você teria de conhecer a falsidade da (s) hipótese (s) cética (s) associada (s) à proposição que você supostamente deveria conhecer. Logo, assumindo o fecho epistêmico, o cético dirá que você deveria conhecer as “proposições-negação” das hipóteses céticas, pois as primeiras são consequências lógicas daquilo que você conhece.

Colocando de outro modo, o cético concluirá seu argumento da seguinte forma – assumindo que você aceite o princípio do fecho epistêmico:

Premissa 1: Suponha que você saiba que está lendo uma tese em epistemologia.

Premissa 2: Suponha que você saiba que, se você está lendo uma tese em epistemologia, então você não é um cérebro numa cuba.

Premissa 3: Ora, acontece que você não sabe – e não tem como saber – que você não é um cérebro numa cuba.

Conclusão: Portanto, por *Modus Tollens*, você não sabe que está lendo uma tese em epistemologia. Para saber isto, você deveria conhecer a falsidade da hipótese cética em questão – o que não é o caso[29].

Deixemos que o cético generalize seu resultado pessimista. Pense em H como qualquer hipótese cética. Pense no símbolo ‘$\neg$’ como um representante formal para a negação. Assim, pense em ‘$\neg H$’ como uma representação simbólica para a negação da hipótese cética H. Em seguida, pense em P como uma proposição qualquer que você conheça. Feito isto, temos agora o “argumento geral da ignorância”

[28]Para que o exemplo fique mais intuitivo, basta relembrar o filme “*The Matrix*”. Sem a ajuda de Morfeu, *Neo* não saberia que ele era prisioneiro da *Matrix*. As percepções que ela criava no cérebro de *Neo* eram tão similares às percepções de uma pessoa em condições normais que, sozinho, *Neo* não seria capaz de descobrir que havia sido enganado desde seu nascimento.

[29]A regra de inferência *Modus Tollens* é, ao mesmo tempo, muito simples e bastante utilizada em lógica proposicional. Ela diz que, se é verdade que P implica Q mas falso que Q, então P também deve ser falsa. Exemplo: se, neste exato momento, estou digitando algo no computador, então devo possuir ao menos uma mão (hipótese). Infelizmente, não possuo nem sequer uma mão (hipótese). Logo, por *Modus Tollens*, não estou digitando algo no computador.

ou *Modus Tollens*:

1. $K_V P$	Hipótese
2. $K_V(P \to \neg H)$	Hipótese
3. $(K_V P \wedge K_V(P \to \neg H)) \to K_V \neg H$	Fecho Epistêmico
4. $\neg K_V \neg H$	Premissa cética
5. $\neg(K_V P \wedge K_V(P \to \neg H))$	3,4 por *Modus Tollens*
6. $\neg K_V P \vee \neg K_V(P \to \neg H)$	5 por De Morgan
7. $\neg K_V P$	2,6 por silogismo disjuntivo

E assim o cético chega à sua conclusão pessimista: não temos conhecimento nem mesmo de proposições simples como "*estou sentado lendo uma tese em epistemologia*". Não sabemos dessas coisas porque as possibilidade de erro – ou hipóteses céticas – estão sempre presentes quando afirmamos conhecer algo ou até mesmo atribuir conhecimento a outros. Algumas dessas possibilidades – como o cenário do cérebro numa cuba – não podem ser eliminadas: é impossível (por hipótese) saber que elas são falsas. Se isso é o caso, o conhecimento parece algo impossível. Eis um grande problema para os epistemólogos, e tudo – como diria alguns – devido a aceitação de (E-CLOS 1)!

Como se pode observar, o princípio (E-CLOS 1) desempenha um papel importante no argumento da ignorância. Não será possível, nele, obter $\neg K_V P$ a menos que você também aceite (E-CLOS 1).

Uma possível solução para esse problema, portanto, seria simplesmente rejeitar (E-CLOS 1). Com isso, se bloquearia o argumento cético antes que ele atingisse o resultado pessimista expresso por '$\neg K_V P$'. Contudo, até mesmo um crítico do fecho – como o próprio Dretske – reconhece a plausibilidade de (E-CLOS 1), e que a ideia de abandoná-lo completamente não é muito popular entre os filósofos:

> Feldman (1999) pensa que abandonar o fecho é "uma das ideias menos plausíveis a ganhar *status* na epistemologia esses últimos anos". DeRose (1995) acha-a intuitivamente "bizarra" ou "abominável". Fumerton (1987) pensa que a falha do fecho é uma "objeção devastante" e Bonjour (1987) um *reductio ad absurdum* de qualquer teoria que a implica ou a adota[30]. (DRETSKE, 2005a, p. 17)

[30]*"Feldman (1999) thinks that abandoning closure is "one of the least plausible ideas to gain currency in epistemology in recent years." DeRose (1995) finds it "intuitively bizzare" or "abominable." Fumerton (1987) thinks the failure of closure is a "devastating objection" and BonJour (1987) a reductio ad absurdum to any theory that implies or embraces it."*

Apesar disto, se existe algo pior do que abandonar o fecho, esse algo seria justamente abraçar o ceticismo. Se abandonar o fecho é a única maneira de escapar à conclusão gerada pelo argumento da ignorância, então a atitude de Dretske de rejeitar o fecho seria justificada. E é precisamente isso que Dretske pretende mostrar (2005a, p. 18):

> [...] se fosse possível demonstrar que a rejeição do fecho não é apenas mais um meio de evitar o ceticismo (muitos filósofos concordariam com isso), mas a única forma de evitar o ceticismo, essa ideia teria um grande peso para filósofos que acham o ceticismo tão "bizarro" ou "abominável" quanto à rejeição do fecho[31].

Vejamos, portanto, a posição de Dretske; investiguemos seus argumentos com maior cautela.

1.5 Dretske: resposta aos céticos e ao fecho epistêmico

Hendricks observa:

> A epistemologia começa com o desagradável ceticismo que surge da possibilidade de um demônio maligno. Qualquer afirmação sobre posse, aquisição ou manutenção de conhecimento, antes que a afirmação cética sobre a impossibilidade de conhecimento seja derrotada, será absurda. Para tirar a epistemologia do chão, é necessário demonstrar que o conhecimento é, de fato, possível[32]... (HENDRICKS, 2006, p. 50)

E é exatamente o que Dretske tenta fazer. Em *Epistemic Operators* (1970), Dretske investiga os "degraus de penetrabilidade" de vários operadores – os epistêmicos estão entre eles. A propriedade de fecho é chamada de "penetrabilidade". Deste modo, se afirmamos que um operador – epistêmico ou não – é "plenamente penetrante", estamos simplesmente afirmando que esse operador é fechado sob implicação lógica (em todos os casos). Similarmente, afirmar que algum operador é

[31] "[...] *if a case could be made for the claim that a rejection of closure was not just a way to avoid skepticism (most philosophers would agree with this) but the only way to avoid skepticism, it should carry weight with philosophers who find skepticism as "bizarre" or "abominable" as the denial of closure.*"

[32] "*Epistemology begins with facing the beastly skepticism that arises from the possibility of an evil demon. Any talk about knowledge possession, acquisition let alone maintenance before skepticism?s claim about the impossibility of knowledge is defeated, is absurd. To get epistemology off the ground it must be demonstrated that knowledge is in fact possible...*"

"não-penetrante" é simplesmente dizer que, até mesmo nos casos mais simples, o fecho não vale para esse operador.

Dretske apresenta um exemplo bem interessante de operador não penetrante. Considere "*é estranho que...*" um operador. Agora, suponha que a seguinte proposição seja verdadeira:

P_7: *Ela perdeu o jogo.*

Novamente, ser-nos-á possível deduzir um número infinito de proposições que são consequências lógicas de P_7. Escolhamos uma proposição particular qualquer:

P_8: *Alguém perdeu o jogo.*

Não será difícil notar que P_7 implica logicamente P_8 . No entanto, como também podemos observar, não podemos deduzir a proposição "*é estranho que alguém tenha perdido o jogo*" da proposição "*é estranho que ela tenha perdido o jogo*". Ora, pode ser estranho que ela tenha perdido o jogo, considerando que ela estava tão preparada... Mas não é, de modo algum, estranho que alguém tenha perdido. A derrota é um fenômeno natural em nossas vidas. Todos perdem pelo menos uma vez em algum jogo. Portanto, com o exemplo acima, Dretske forneceu um caso de operador não-penetrante.

Mas o que dizer do operador epistêmico "*conhece que...*" (ou "*sabe que...*")? Para Dretske, "*conhece que...*" é um exemplo do que ele chama de "operador semi penetrante". Obviamente, esse operador não pode ser não-penetrante, pois seu degrau de penetrabilidade é maior do que o de operadores como "*é estranho que...*". Há um exemplo muito simples para ilustrar isso. Suponha que você conheça a seguinte proposição:

P_9: *Smith e Jones foram convidados para a festa.*

Certamente, uma consequência lógica de P_9 seria:

P_{10}: *Smith foi convidado para a festa.*

Parece-nos aceitável que, se um agente conhece P_9, ele também conhece P_{10}. Deste modo, como um resultado trivial, "*conhece que...*" – que é um operador semi penetrante – penetra mais profundamente do que operadores não-penetrantes.

Contudo, o interesse de Dretske também é demonstrar que o operador "*conhece que...*" não é plenamente penetrante: há casos em que o fecho não vale para esse operador. Assim, argumenta Dretske, há casos que em (i) "*S conhece P*", (ii) "*S Conhece* $(P \to Q)$" mas (iii) "*S não conhece Q*". E, como vimos anteriormente, isso ocorre exatamente nos casos em que Q é a negação de alguma hipótese cética – como, por exemplo, "*eu não sou um cérebro numa cuba*".

Observamos, até agora, que o fecho epistêmico dá espaço ao ceticismo através do argumento da ignorância. Como Dretske está ciente disso, sua estratégia é bloquear o referido argumento através da rejeição do princípio que o torna viável – o próprio fecho. Com a rejeição do fecho, o cético não obterá seu argumento, dado que não terá à disposição todas as premissas necessárias para sua derivação.

Entretanto, para uma plena compreensão do argumento de Dretske, faz-se necessário entender o que ele quis dizer com o termo "consequências de contraste":

> Suponha que asserimos que x é A. Considere algum predicado, 'B', que é incompatível com A, de modo que nada pode ser, ao mesmo tempo, A e B. Segue-se, portanto, do fato de que x é A, que x não é B. Além disso, se conjuntamos B com qualquer outro predicado Q, seguir-se-á, do fato de que x é A, que x é não-(B e Q). Chamarei esse tipo de consequência de consequência de contraste, e estou interessado em um subconjunto dessa classe: pois acredito que as objeções céticas mais convincentes às nossas afirmações de conhecimento do dia a dia exploram um conjunto particular dessas consequências de contraste[33]. (DRETSKE, 1970, p. 1015)

Como exemplo, considere a proposição P_1: "Estou lendo uma tese em Epistemologia". Para cada proposição – seguindo o raciocínio de Dretske – podemos associar uma outra que é consequência de contraste da primeira. Deste modo, podemos associar a P_1 a seguinte consequência de contraste:

P_4: *Não estou jogando videogame.*

Agora, suponha que você saiba que P_1 seja o caso. Suponha também que você saiba que, se está lendo uma tese em epistemologia, então você não está jo-

[33] *"Suppose we assert that x is A. Consider some predicate, 'B', which is incompatible with A, such that nothing can be both A and B. It then follows from the fact that x is A that x is not B. Furthermore, if we conjoin B with any other predicate Q, it follows from the fact that x is A that x is not-(B and Q). I shall call this type of consequence a contrast consequence, and I'm interested in a particular subset of these: for I believe the most telling skeptical objections to our ordinary knowledge claims exploit a particular set of these contrast consequences."*

gando videogame. Podemos concluir disso, pelo fecho epistêmico, que você sabe que não está jogando videogame.

Se você observar de perto esse exemplo, tudo parecerá estar em ordem. Na verdade, está. Para Dretske, ao considerarmos a proposição "*não estou jogando videogame*", parece que estamos pressupondo o "estar jogando videogame" como uma "possibilidade relevante". Porém, considere agora alguns argumentos céticos que contêm "estranhas" consequências de contraste:

Argumento cético A1.

Sejam as proposições:

P: A parede é vermelha.

Q: A parede é branca.

R: A parede está sendo engenhosamente iluminada para parecer vermelha.

Como se pode ver, P implica não-Q e não-R e, portanto, implica não-(Q e R). Suponha que você conhece P e também estas implicações. Assim, suponha que você conhece tanto P quanto P implica não-(Q e R). Pelo fecho epistêmico, você também deve conhecer não-(Q e R). Mas você realmente sabe que não-(Q e R) é o caso? Ora, considerando que você não tem como saber não-(Q e R) – porque não tem como saber não-R – conclui-se que você não sabe que P, a saber, que a parede é vermelha[34].

Argumento cético A2.

Sejam as proposições:

P: Aquelas são zebras.

Q: Aquelas são mulas.

R: Aquelas são mulas engenhosamente disfarçadas para perecerem zebras.

[34]É importante observar que, de uma perspectiva mais rigorosa, a proposição "não-(Q e R)" não é uma consequência lógica direta de P, não-Q e não-R. O argumento de Dretske está, portanto, incompleto. Uma prova rigorosa do argumento A1 seria mais ou menos assim: para redução ao absurdo, podemos supor que (Q e R) é verdadeiro. Como (Q e R) é uma conjunção, inferimos Q. Porém, sabemos tanto que P é o caso como também que P implica não-Q. Assim, por eliminação da implicação (ou *Modus Ponens*), inferimos não-Q. Mas aí encontramos uma contradição entre Q e não-Q. Assim, nossa hipótese (Q e R) deve ser falsa. Portanto, por redução ao absurdo, não-(Q e R) deve ser verdadeira. As mesmas considerações se aplicam ao argumento A2.

Novamente, P implica não-(Q e R). Suponha que você conhece P, bem como a implicação $P \rightarrow$ não-(Q e R). Pelo fecho epistêmico, você também deveria saber que não-(Q e R). Contudo, você não sabe disso. Portanto, você não conhece P.

Os argumentos céticos A1 e A2 utilizam consequências de contraste similares. A consequência de contraste da forma não-(Q e R), em ambos os casos, parece bastante improvável. De fato, quando você observa uma parede vermelha, você geralmente não considera a possibilidade de estar sendo enganado por falsa iluminação ou, talvez, demônios malignos. Em outras palavras: essas não são "possibilidades relevantes" para você. O mesmo se aplica ao caso da zebra.

No entanto, se você prestar mais atenção aos dois casos, verá que essas possibilidades remotas ainda não deixam de ser possibilidades. Deste modo, a pergunta "E se...?" permanece, incomodando você. É claro que você pode e – provavelmente – insistirá: "Mas aquelas são zebras, eu tenho certeza!" Porém, o cético pode replicar: "E se elas fossem, na realidade, mulas engenhosamente disfarçadas pelas autoridades do zoológico? Você está certo, sem qualquer sombra de dúvida, que isso não seja o caso?"

Dretske vê uma saída desse problema. Ele argumenta que, sem qualquer dúvida, você conhece que "a parede é vermelha" e que "aqueles animais são zebras". Para que você conheça as respectivas proposições, não é necessário que também conheça a falsidade das hipóteses céticas associadas a elas, pois o operador de conhecimento não penetra nas consequências de contraste comumente apresentadas pelo cético. Em outras palavras, o fecho epistêmico não vale nos casos em que a consequência lógica do conhecimento de alguém é a negação de alguma hipótese cética. Assim, se ele – o fecho – não vale, você não é obrigado a conhecer todas as consequências lógicas daquilo que você conhece. Nos casos acima, que envolvem negações de hipóteses céticas como consequência de contraste, mesmo ao se assumir que você conhece P, não podemos requerer de você que também conheça a falsidade de (Q e R). Logo, sem qualquer receio, você pode afirmar: "Eu não sei se (Q e R) ou não-(Q e R), e daí?" Enfim, seu conhecimento de proposições não-céticas está seguro novamente.

Depois de mais de trinta anos (2005), a posição de Dretske com relação ao fecho epistêmico continua a mesma. Por quê? Na verdade, essa sua postura sobre o fecho é uma consequência necessária de sua concepção acerca da noção de conhecimento (DRETSKE, 1971). Em sua opinião, conhecer uma proposição P, por exemplo, é possuir razões conclusivas para P. O que isso significa? Dretske define o conceito de razões conclusivas contrafactualmente:

CR: "R é uma razão conclusiva para P" é equivalente a "se P fosse falsa, R também seria falsa" (ou: "se P não fosse o caso, R também não seria o caso")[35].

Agora, considere a proposição P_1 mais uma vez:

P_1: *Estou lendo uma tese em Epistemologia.*

Se você conhece P_1, então – segundo Dretske – você tem razões conclusivas para P_1. Seja R uma abreviação para esse conjunto de razões conclusivas[36]. Assim, se você conhece P_1, é porque você possui R. A condição *CR* diz que, se P_1 fosse falsa, você não teria R. Suponhamos, portanto, que P_1 seja verdadeira – isto é, que você realmente esteja lendo uma tese em Epistemologia. E agora? Você possui razões conclusivas para isto? Para alcançar o resultado desejado, devemos verificar se a proposição "*O agente S* [você] *está agora lendo uma tese em Epistemologia*" satisfaz a condição *CR*: Se P_1 fosse falsa, o agente S não teria R. As evidências para P_1 são muitas. Você pode ver o papel bem à sua frente. Você também pode ver as letras, bem como compreender as palavras utilizadas pelo autor. Além disso, você pode compreender todos os conceitos relevantes, seus argumentos, razões etc. Se a tese for impressa, também é possível tocá-la com suas mãos. Ora, se você não estivesse lendo esta tese, isto é, se P_1 fosse falsa, então você não teria todas essas evidências para a proposição P_1. Portanto, se P_1 fosse falsa, você não teria R. Como se pode ver, a condição *CR* é satisfeita. Você sabe que está lendo uma tese em epistemologia.

Mas o que dizer de proposições que envolvem hipóteses céticas? Não é difícil notar que você nunca terá razões conclusivas para proposições como "*Não sou um cérebro numa cuba*" ou "*Não sou uma marionete do gênio maligno cartesiano*". Vejamos o porquê:

Considere a proposição:

P_2: *Não sou um cérebro numa cuba.*

Suas razões para acreditar em P_2 também são muitas. Neste exato momento, você está lendo uma tese em epistemologia (cérebros não leem). Além disso,

[35] *CR* é uma abreviação para "*conclusive reasons*."
[36] R pode ser pensado como o conjunto de todas as suas evidências para a proposição em questão.

você está segurando a tese com suas próprias mãos (as mesmas mãos que você pode sentir); você pode ver o papel, tocá-lo, sentir o seu cheiro etc. Ora, cérebros não fazem todas essas coisas! A condição *CR* diz que, se P_2 fosse falsa, você não teria todas as evidências que você tem para acreditar na proposição P_2. Assim, *CR* diz que, se você fosse um cérebro numa cuba, você não veria suas mãos, a tese e assim por diante.

Mas será que isso é mesmo verdade? Dretske afirma que não: o cientista maluco, que aprisionou o seu cérebro, teria o poder de enganá-lo acerca de quase tudo. Ele poderia fazer o seu cérebro pensar que é uma pessoa de verdade (com mãos, juntas etc.) quando de fato não o é. Se essa situação fosse satisfeita, você teria R mesmo que P_2 fosse falsa. Mas isso contradiz a condição *CR*. Logo, você não tem razões conclusivas para P_2 e, portanto, não sabe que P_2 é o caso. O fecho epistêmico, neste caso particular, deve ser falso porque:

1. Você pode possuir razões conclusivas para P_1;
2. Você pode possuir razões conclusivas para $P_1 \rightarrow P_2$;
3. Você não pode possuir razões conclusivas para P_2.

Como vimos, a teoria da razões conclusivas sugere a rejeição do fecho epistêmico em alguns casos particulares. Além disso, outras razões podem ser encontradas para sustentar esta posição (DRETSKE, 2005a e 2005b). Uma delas é a falha do fecho no que se refere à transmissão de garantia evidencial. Mesmo que possua evidências perceptuais para a proposição que expressa o fato de eu estar escrevendo uma tese em epistemologia neste exato momento, isso não me permite concluir que tenho evidências perceptuais para a proposição logicamente implicada por esse fato, a saber, que o mundo material existe. Deste modo, apesar de possuir uma experiência perceptual do computador que, neste exato momento, está diante de meus olhos, eu não tenho – necessariamente – o conhecimento de que o mundo externo existe – mesmo que eu saiba que o primeiro fato implica o segundo. Dretske nos mostra que um meio de descobrir P (há um computador diante de mim) não é, necessariamente, o mesmo meio de descobrir Q (o mundo externo existe) – mesmo que eu esteja ciente (ou saiba) que P implica Q. Logo, o seguinte argumento é inválido (DRETSKE, 2005a, p. 16):

R é uma razão para S acreditar em P.

S sabe que P implica Q.

$\therefore R$ é uma razão para S acreditar em Q.

Apesar disso, Dretske reconhece que a falha na transmissibilidade de evidência, por si só, não invalida o fecho epistêmico:

> A não-transmissibilidade, por si só, não implica na falha do fecho pois, como nosso exemplo do vinho ilustra, mesmo que as razões de S para acreditar em P não sejam transmitidas a uma consequência conhecida, Q, ainda é possível que S deva conhecer Q (talvez baseado em outras razões) para que possa conhecer P[37]. (DRETSKE, 2005a, p. 15)

Porém, argumenta, "uma vez apreciada a falha extensiva da transmissão evidencial, a falha do fecho é, se não obrigatória, ao menos mais fácil de engolir." (DRETSKE, 2005a, p. 15)

Para completar a tarefa de invalidação do fecho, Dretske fará uso de suas famosas "proposições *heavyweight*". Mas o que são elas? Qual sua significância? Essas proposições são de grande importância, pois elas são os pilares de sua teoria contra o fecho. Hawthorne, adversário de Dretske, as define do seguinte modo:

> Seja P uma "proposição *heavyweight*" exatamente quando todos possuímos uma inclinação forte para pensar que P não seja o tipo de coisa que se possa conhecer apenas através do exercício da razão, e também que P não seja o tipo de coisa que se possa conhecer pela utilização das faculdades perceptuais (mesmo quando estas últimas sejam auxiliadas pela razão)[38]. (HAWTHORNE, 2005, p. 33)

Então, basicamente, uma proposição *heavyweight* é aquela que não é facilmente cognoscível. Observa-se, no entanto, que a definição não é precisa. Quando uma proposição deve ser considerada *heavyweight*? Eis uma questão difícil. Todavia, ainda é possível apresentar algumas proposições que seriam, provavelmente, consideradas *heavyweight*. Tome, como exemplo, a proposição P_2. A proposição "*Não sou um cérebro numa cuba*" é do tipo que pode ser conhecida pela percepção? Como sabemos, muitas pessoas diriam que não. Na verdade, eu também diria que não. Filosoficamente falando, a proposição em questão não é do tipo que pode ser

[37]*"Non-transmissibility does not itself imply the failure of closure since, as our wine example illustrates, even when S's reasons for believing P do not transmit to a known consequence, Q, it may be that S must still know Q (perhaps on the basis of other reasons) in order to know P".*

[38]*"Let P be a "heavyweight proposition" just in case we all have some strong inclination to think that P is not the sort of thing that one can know by the exercise of reason alone and also that P is not the sort of thing that one can know by use of one's perceptual faculties (even aided by reason)."*

conhecida apenas com base em nossas próprias percepções sensoriais. Mas será que, por acaso, não existe um outro meio de descobrir que P_2 é verdadeira? Alguns diriam que sim, outros diriam que não. O fato é que P_2 não é uma proposição fácil de provar – filosoficamente, é claro. Se por acaso o fosse, o desafio cético talvez não existisse. Infelizmente, todos sabemos que o desafio cético existe e continua a exigir atenção da epistemologia contemporânea. O ceticismo não é tão fácil de refutar (e talvez seja até mesmo irrefutável).

1.5.1 O efeito das proposições *heavyweight*

Um detalhe interessante sobre as proposições *heavyweight* é justamente o de que seu número é infinito. Através de um simples raciocínio, temos o seguinte resultado:

Para cada proposição P existe uma outra proposição, Q, de modo que Q pode ser associada a P como uma consequência lógica de P, e Q é uma proposição *heavyweight*. Exemplo:

- *P: Estou vivo* (proposição verdadeira escolhida arbitrariamente).
- *Q: Não estou enterrado no cemitério local*;
- Q_1: *Não estou no céu*;
- Q_2: *Não estou no Hades*;
- ...
- Q_∞.

Vimos que a percepção não é uma fonte adequada de conhecimento de proposições *heavyweight*. Seria possível, talvez, conhecê-las de alguma outra forma – ou por outra fonte. Infelizmente, o exemplo acima apoia a posição de Dretske, pois todos os meios de conhecer proposições *heavyweight* "ou falham em alcançar essas implicações *heavyweight* ou geram suas próprias implicações *heavyweight*" (DRETSKE, 2005a, p. 33). Outros meios de conhecer Q seriam através do testemunho ou da memória, ou algo do tipo. Infelizmente, é possível gerar implicações *heavyweight* para cada fonte de conhecimento que possamos pensar.

O argumento baseado em proposições *heavyweight* é, sem dúvida, atraente. Aqueles que defendem o fecho deveriam fornecer alguma forma de refutação para ele. Mas como seria tal refutação? O que ela deveria dizer? Deveria ela afirmar que as proposições *heavyweight* não existem? Ou será que deveria, diferentemente de Dretske, afirmar que elas são facilmente cognoscíveis? Será que tal refutação deveria afirmar que a linha que divide a proposições *heavyweight* das proposições *lightweight* ou *middleweight* não estaria bem traçada? Alguns defensores do fecho, incluindo entre eles o próprio Hawthorne, se confundem um pouco quando respondem ao desafio de Dretske. Para ver o porquê disso, e também para analisar as respostas de Hawthorne (2005) às questões que acabamos de levantar, nos debruçemos agora sobre a posição dele com relação a Dretske e ao fecho epistêmico.

1.6 Resposta de Hawthorne a Dretske

O primeiro detalhe importante sobre a abordagem de Hawthorne é a sua própria versão preferida do fecho epistêmico. Baseando-se na ideia de Williamson de que a dedução é um modo de expandir o conhecimento (WILLIAMSON, 2000, p. 117), Hawthorne apresenta uma versão atraente de fecho:

(E-CLOS-H): Se S conhece P e deduz competentemente Q de P, vindo a acreditar em Q enquanto retém o conhecimento de P, então S vem a conhecer Q.

Apesar do fato de esta versão particular de fecho não ser o alvo original de Dretske, todos os seus argumentos contra (E-CLOS 1) também se aplicam a (E-CLOS H). Como sabemos, as pessoas não saem por aí acreditando seriamente que são cérebros em cubas ou marionetes do gênio maligno cartesiano. Isso ocorre porque elas sabem, por exemplo, que elas possuem corpos, e que 'possuir um corpo' implica em 'não ser um cérebro numa cuba'. Deste modo: (i). Elas conhecem *P* (elas têm corpos); (ii). Elas sabem que *P* (ter um corpo) implica *Q* (não ser um cérebro numa cuba); (iii). Elas acreditam em *Q* (elas não são cérebros em cubas), enquanto retêm conhecimento das premissas (i) e (ii). (iv). Elas vêm a saber que não são cérebros em cubas.

Poderíamos nos perguntar se há alguma coisa errada com esse raciocínio. Hawthorne diria que não, já Dretske diria que sim. Por que Dretske iria insistir em sua posição? A resposta é simples: a conclusão do argumento é a negação de uma hipótese cética. Ou seja, uma proposição *heavyweight* ainda aparece como conclusão do referido argumento.

A seguir, irei apresentar os argumentos de Hawthorne contra a posição de Dretske (juntamente com algumas objeções minhas a Hawthorne), com foco central nas proposições *heavyweight*. Contudo, deixarei as principais objeções para a próxima seção, em que as intuições de McBride (2009) sobre assunto serão analisadas. Por enquanto, voltemos nossa atenção para um argumento muito simples, baseado no "princípio da equivalência":

(EQP) Princípio da equivalência: Se S conhece a priori que P é equivalente a Q, conhece P e deduz Q de P, retendo o conhecimento de P, então S conhece Q[39].

Argumento de Hawthorne: Dretske está comprometido com a rejeição da distribuição. Para isso, suponha que um agente qualquer, *S*, conhece *P*: "A taça está cheia de vinho". Seja *Q* a seguinte proposição: "A taça está cheia de um líquido colorido artificialmente para parecer com o vinho." Seguindo o raciocínio, a proposição *P* é "equivalente *a priori*" a "*P* e não-*Q*". Daí, pelo princípio da equivalência, *S* conhece a respectiva conjunção. Agora, pelo princípio da distribuição, se *S* conhece a referida conjunção, *S* conhece os constituintes da mesma. Isto é, se *S* conhece '*P* e não-*Q*', *S* também deve conhecer não-*Q*. Porém, segundo Dretske, *S* não pode saber disso. Conclusão: se a posição de Dretske fosse aceita, o princípio da distribuição deveria ser rejeitado.

Observação crítica: Primeiramente, os *insights* iniciais de Dretske acerca da distribuição são favoráveis, não o contrário (DRETSKE, 1970, p. 1009). Isto é, em "*Epistemic Operators*" (1970), Dretske aceita o princípio da distribuição. Além disso, se por acaso for insistido que a abordagem das razões conclusivas força a rejeição do princípio da distribuição, então será possível fornecer uma implementação não *ad hoc* da definição de razões conclusivas que será capaz de bloquear a conjunção '*P* e não-*Q*' (MCBRIDE, 2009), impedindo o argumento de Hawthorne através da rejeição de premissas[40]. O resultado é o seguinte: o desafio proposto por Dretske ainda não foi solucionado.

Mas isso é apenas o começo. Hawthorne realmente oferece diferentes argumentos contra a posição de Dretske. Como vimos, a noção de proposição *heavyweight* é central na visão crítica de Dretske. Deste modo, se ela estiver equivocada, a "estratégia *heavyweight*" pode nos levar a lugar nenhum. Ou seja, a

[39](EQP) é uma abreviação para *equivalence principle*.

[40]No presente caso, com a rejeição da premissa de que *S* conhece a conjunção entre *P* e não-*Q*.

aceitação de proposições *heavyweight* pode ser um beco sem saída. Hawthorne escreve:

> Parece evidente que a abordagem de Dretske foi projetada para fornecer a conclusão de que é possível conhecer proposições comuns e suas consequências "não manifestamente *heavyweight*", enquanto se permanece ignorante de suas consequências manifestamente *heavyweight*. Mas a mesma não chega nem perto de fornecer esse resultado. Acontece que todos nós não raramente temos razões conclusivas para proposições manifestamente *heavyweight* (e que, nesses casos, Dretske realmente não oferece qualquer barreira efetiva ao fato de conhecermos essas proposições) e, não raramente, carecemos de razões conclusivas para consequências *a priori* de proposições conhecidas, mesmo que tais consequências não sejam manifestamente *heavyweight*[41]. (HAWTHORNE, 2005, p. 35)

Para estabelecer isto, Hawthorne precisa apresentar um caso em que: (i) Um agente '*S*' tenha razões conclusivas para uma proposição manifestamente *heavyweight* Q. (ii) Um agente '*S*' careça de razões conclusivas para Q, sendo Q uma proposição não-*heavyweight*. Vejamos como ele faz as duas coisas (HAWTHORNE, 2005, p. 35).

1.6.1 Proposições manifestamente *heavyweight* com razões conclusivas

Caso de Hawthorne número 1 (o caso da dor de cabeça)

Considere a proposição P: "Eu não sou um cérebro numa cuba". Ora, posso me perguntar se sei ou não se P é verdadeira. De acordo com Dretske, eu não tenho como saber isso, pois careço de razões conclusivas para P. Isto é: se P fosse falsa, eu ainda teria R (um conjunto de evidências de apoio para P). Agora, considere a proposição Q: "Eu tenho uma dor de cabeça e eu não sou um cérebro numa cuba". Parece que tenho razões conclusivas para Q. Considerando que meu conjunto de evidências inclui minha dor de cabeça, temos que: se Q fosse falsa, eu não teria R. Contudo, Q é uma proposição *heavyweight*. Logo, eu tenho razões conclusivas para uma proposição *heavyweight*.

[41] *"It seems evident that Dretske's account is designed to deliver the conclusion that one knows ordinary propositions and their non-manifestly heavyweight consequences, while remaining ignorant of their manifestly heavyweight consequences. But it does not come close to delivering that result. It turns out that we all too often have conclusive reasons for manifestly heavyweight propositions (in which case Dretske has not in fact provided an effective barrier to knowing such propositions), and all too often lack conclusive reasons for a priori consequences of known propositions, even though those consequences are not manifestly heavyweight."*

Se isto estiver correto, a posição de Dretske possui algum defeito. Dada a possibilidade de se conhecer proposições *heavyweight*, não será necessário restringir o fecho apenas às proposições *light* e *middleweight*.

<u>Caso de Hawthorne número 2</u> (o caso "parece um pássaro")

Suponha que você esteja num zoológico. Suponha ainda que você esteja vendo um pássaro voando em uma gaiola próxima a você. Com base em suas evidências perceptuais, você conclui que existe um pássaro na gaiola. "Neste momento – você raciocina – o pássaro que eu vejo não é um objeto inanimado, engenhosamente disfarçado para parecer com um pássaro de verdade." É evidente que a proposição "neste momento, o pássaro que eu vejo não é um objeto inanimado, engenhosamente disfarçado para parecer como um pássaro de verdade" é *heavyweight*. Seja Q a abreviação para esta proposição. Agora, há um detalhe importante: se alguém fosse fazer algum objeto inanimado parecer com um objeto animado, esse alguém iria escolher algo muito mais fácil de imitar; esse alguém escolheria, por exemplo, fazer uma imitação de tartaruga. Deste modo, nos mundos mais próximos em que há um objeto inanimado na gaiola (engenhosamente disfarçado para enganá-lo), não há qualquer coisa parecida com um pássaro. Ou seja, nos mundos mais próximos em que existe um objeto inanimado engenhosamente disfarçado, tal objeto não pareceria com um pássaro, e sim com uma tartaruga – que, plausivelmente, seria mais fácil de imitar. Mas repare que agora temos: "se Q fosse falsa, você não teria R"; isto é, se não existisse um pássaro na gaiola, você não o veria. Logo, neste caso, você tem razões conclusivas para Q – que é *heavyweight*.

<u>Caso de Hawthorne número 3</u> (o "caso do biscoito")

Suponha uma situação em que há um biscoito diante de você. Devido a suas experiências perceptuais, você constrói a crença de que existe um objeto real que está, grosso modo, a um metro e meio de distância de você[42]. A proposição "existe um objeto real que está a um metro e meio de distância de mim" pode ser caracterizada como *heavyweight*. Novamente, seja Q a abreviação para esta proposição. Agora, como no caso anterior, se Q fosse falsa, você não teria R: nos mundos mais próximos, se não houvesse um objeto real a um metro e meio de você, você não acreditaria existir um objeto real a um metro e meio de você – aquele lugar não estaria sendo ocupado por qualquer coisa e, por conta disso, você não acreditaria

[42]O termo original em inglês é *mind-independent object*. Com isso, o autor sugere um objeto que possui realidade objetiva, isto é, existe fora da mente do agente em questão; em outras palavras, não precisa do agente para existir, não é uma criação de sua mente.

haver qualquer objeto ali. Logo, mais uma vez, você tem razões conclusivas para Q – que é *heavyweight*.

Devemos observar que, se os casos 1, 2 e 3 estiverem corretos, a tese defendida por Hawthorne constitui uma objeção séria à teoria das proposições *heavyweight*, de Dretske. Entretanto, vejamos com mais cautela os desdobramentos da argumentação oferecida por Hawthorne.

1.6.2 Implicações dos casos de Hawthorne

Como podemos observar, os três casos apresentados acima pretendem demonstrar algo bastante simples: é possível possuir razões conclusivas para proposições *heavyweight*. Se isso for verdade, então não há nada de errado com o fecho epistêmico. Ou seja, se podemos ter razões conclusivas para proposições *heavyweight* – ou em outras palavras, se proposições *heavyweight* são cognoscíveis – então o fecho epistêmico não está restrito apenas a proposições *lightweight* e *middleweight*. A abordagem de Dretske é irrelevante.

No entanto, o argumento de Hawthorne parece estar incompleto. A menos que ele possa provar que podemos ter razões conclusivas para TODAS as proposições *heavyweight*, seu argumento não funcionará como desejado. Pois, se acontecer de existir pelo menos uma proposição que seja, ao mesmo tempo, *heavyweight* e incognoscível, isso será suficiente para manter Dretske no jogo. Isso ocorre porque a tese de Dretske é particular e negativa. Ele pensa o fecho como sendo inválido. Para sustentar sua posição, ele precisa apenas encontrar algum caso particular em que o fecho não funcione. Por outro lado, Hawthorne deve demonstrar que toda proposição supostamente *heavyweight* de Dretske não é incognoscível, já que é possível encontrar razões conclusivas para cada uma delas.

Com isso, fica fácil perceber a insuficiência do argumento de Hawthorne. Mesmo se os três casos apresentados por ele estiverem corretos, ainda assim eles serão todos particulares. Porém, considerando que seu argumento requer um resultado geral, as premissas apresentadas (os três casos em que proposições *heavyweight* são cognoscíveis) não poderão garantir a conclusão. Em outras palavras, o argumento oferecido por Hawthorne não garante a generalidade pretendida para a rejeição completa da teoria de Dretske. Mesmo que listemos infinitos casos particulares em que proposições *heavyweight* são cognoscíveis, ainda assim será possível encontrar uma proposição *heavyweight* incognoscível. Hawthorne deveria fornecer algum mecanismo que garantisse a generalidade de que seu argumento necessita para funcionar. Porém, ele não faz isso. Logo, a posição – correta ou incorreta – de Dretske continua firme. O conhecido "desafio de Dretske" não foi satisfatoriamente solucionado por Hawthorne.

Apesar disso, Hawthorne insiste na segunda parte de seu argumento. Vejamos como.

1.6.3 Proposições não-manifestamente *heavyweight* sem razões conclusivas

Os casos seguintes pretendem mostrar que, em algumas circunstâncias, carecemos de razões conclusivas para proposições não-*heavyweight*. Se isso puder ser estabelecido, ficará demonstrado que toda a iniciativa da abordagem das proposições *heavyweight* não traz qualquer resultado relevante para a discussão do fecho – assim pensa Hawthorne. O raciocínio por traz desta estratégia é bastante simples:

(1) A abordagem *heavyweight* "...foi projetada para fornecer a conclusão de que é possível conhecer proposições comuns e suas consequências "não manifestamente *heavyweight*", enquanto se permanece ignorante de suas consequências manifestamente *heavyweight*"; (HAWTHORNE, 2005, p. 35)

(2) Podemos possuir razões conclusivas para proposições *heavyweight*;

(3) Carecemos de razões conclusivas para algumas proposições não-*heavyweight*;

(4) A "abordagem *heavyweight*" não fornece o resultado que ela mesmo estabeleceu em suas metas.

(5) Logo, a "abordagem *heavyweight*" está errada.

Até agora, seguimos Hawthorne e tomamos (1) como premissa. Os casos 1, 2 e 3 de Hawthorne – sobre a cognoscibilidade de proposições *heavyweight* – "estabeleceram" (2). Agora, para terminar, (3) é requerido[43]. Para estabelecer (3), Hawthorne utiliza alguns casos bem humorados. Vejamos o conhecido "caso do salmão".

[43]O passo (4) segue dos passos (2) e (3), na medida em que a abordagem inicial de Dretske pretende estabelecer duas coisas: (i) não é possível encontrar razões conclusivas para proposições *heavyweight* e (ii), sempre temos razões conclusivas para proposições não-*heavyweight*. Por conta disso, para que o argumento cumpra sua função, o mais importante é estabelecer (2) e (3). Os passos restantes são consequências dessas duas teses.

Caso de Hawthorne número 4 (o "caso do salmão")

Suponha que você tenha comido na janta uma certa quantidade de salmão. Já que você não é guloso, você forma a crença de que comeu menos de 14 quilos de salmão na janta. Seja Q a abreviação para a seguinte proposição: "*eu comi menos do que 14 quilos de salmão no jantar*". Assim, depois do jantar, você forma a crença de que Q é o caso. Em princípio, não há qualquer coisa anormal com a proposição Q, de modo que Q não deve ser considerada *heavyweight*. Considere Q, portanto, como uma proposição comum – uma proposição não-*heavyweight*. Deste modo, após o jantar, você forma uma crença em uma proposição não-*heavyweight*. A pergunta é: você possui razões conclusivas para a proposição Q? Ao que parece, não... Vejamos o porquê. Suponha, adicionalmente, que o ato de comer mais do que 14 quilos de salmão o induziria a ter alucinações sobre haver comido menos de 1 quilo de salmão. Agora, lembre-se da definição de razões conclusivas: "Se Q não fosse o caso, R não seria o caso."

Suponha, por hipótese, que a proposição "*eu comi menos do que 14 quilos de salmão no jantar*" seja falsa. Logo, você comeu 14 quilos ou mais de salmão no jantar. Mas isso causaria alucinações em você, de modo que você continuaria acreditando que comeu menos do que 14 quilos de salmão no jantar! Logo, se Q fosse falsa, R ainda seria o caso! Conclusão: apesar de Q não ser *heavyweight*, você carece de razões conclusivas para Q.

1.7 Resposta de McBride a Hawthorne

Felizmente, para Dretske, é possível fazer uma implementação não *ad hoc* na definição de razões conclusivas. Tal implementação é fornecida por McBride (2009), de um modo inteligível. Segundo ele, a tese de Dretske deveria ser implementada com a seguinte condição (MCBRIDE, 2009, p. 123):

(SUPP): se P é uma proposição conjuntiva, R não seria verdadeira caso cada um dos conjuntos de P, tomados separadamente, não o fosse[44].

Podemos nos perguntar se esta condição é apenas *ad hoc*. No entanto, McBride nos mostra que ela não o é. Suponha, por exemplo, que você tenha uma dor

[44]Ou seja, se P é uma proposição do tipo $(A \wedge B)$, sendo A e B também proposições, R não seria verdadeira caso A ou B não o fosse; para que R seja verdadeira, A e B devem, igualmente, ser verdadeiras.

de cabeça. Dado isso, está claro que você tem razões conclusivas para a proposição:

P: Eu tenho uma dor de cabeça.

Suas razões para acreditar nesta proposição incluem sua própria dor de cabeça. Agora, se (SUPP) não fosse aceita, como consequência você teria razões conclusivas para a seguinte proposição:

Q: Eu tenho uma dor de cabeça e eu possuo todos os meus membros.

Isso porque os mundos em que é falso que você tem uma dor de cabeça são mais próximos daqueles em que você não possui algum membro. Precisamente nestes primeiros mundos, a condição *CR* é satisfeita:

Se Q fosse falsa, eu não teria R.

Isto é, se você não estivesse com dor de cabeça, você não teria razões para acreditar estar com dor de cabeça. Deste modo, se (SUPP) não for aceita, você pode ter razões conclusivas para *Q* simplesmente se baseando no fato de que você tem razões conclusivas para *P*, e isto não é verdade.

Por conta disto, sustenta McBride, somos forçados a admitir (SUPP) como condição necessária à definição de razões conclusivas. Com (SUPP), é fácil compreender porque o caso número 1 de Hawthorne não apresenta qualquer ameaça à teoria das razões conclusivas.

Resposta de McBride ao 'caso 1' de Hawthorne – utilizando (SUPP)

McBride apresenta, grosso modo, o seguinte argumento:

Seja *S* um agente qualquer. O agente *S* conhece a proposição *P*: "*Eu tenho uma dor de cabeça e eu não sou uma cérebro numa cuba*" somente se[45]:

(1). *S* possui razões conclusivas para "*eu tenho tenho uma dor de cabeça*;"

(2). *S* possui razões conclusivas para "*eu não sou um cérebro numa cuba*."

Ora, *S* carece de razões conclusivas para "*eu não sou um cérebro numa cuba*". Logo, por (SUPP), *S* carece de razões conclusivas para a proposição "*eu tenho uma dor de cabeça e eu não sou um cérebro numa cuba*". Mas a proposição *P* – que é uma conjunção – é *heavyweight*. Conclusão: *S* carece de razões conclusivas para uma proposição *heavyweight* – isto é, para *P* que, como vimos, é *heavyweight*.

[45]O pronome "Eu" se refere ao agente em questão, isto é, a *S*.

Segundo McBride, os casos 2 e 3 também têm uma solução simples. Desta vez, o apelo a (SUPP) nem mesmo é necessário. Isso acontece porque, como McBride coloca, ambos os casos não são instâncias genuínas de proposições *heavyweight*. Comecemos pelo caso 2.

Resposta de McBride ao 'caso 2' de Hawthorne – sem (SUPP)

Antes de tudo, McBride observa que a proposição "*neste momento, o pássaro que eu vejo não é um objeto inanimado, engenhosamente disfarçado para parecer como um pássaro de verdade*" não é *heavyweight*. A resposta para isso é simples: Hawthorne assumiu que, nos mundos mais próximos em que há de fato uma imitação, tal imitação não seria de um pássaro, mas de algo muito mais fácil de imitar – uma tartaruga, talvez. Deste modo, a possibilidade de um objeto inanimado que seja parecido com um pássaro é muito remota para ser considerada como relevante. Se isso for o caso, eu posso conhecer a proposição em questão apenas com base na percepção do pássaro voando pela gaiola. Porém – devemos nos lembrar – se eu posso fazer isto, então a referida proposição não será *heavyweight*. Proposições *heavyweight* não podem ser conhecidas apenas com base em percepções.

Por outro lado, se a possibilidade de um objeto parecido com um pássaro for considerada seriamente, os mundos mais próximos irão permitir objetos inanimados parecidos com pássaros. Daí, se esta possibilidade for levada a sério, a proposição em questão deverá contar como *heavyweight*. Nestes mundos, um agente acreditaria na respectiva proposição ainda que a mesma fosse falsa. Em outras palavras, o agente em questão careceria de razões conclusivas para a referida proposição.

A resposta de McBride ao 'caso 3' de Hawthorne – sem (SUPP)

Novamente, neste caso, McBride mostra que Hawthorne falha em fornecer uma proposição genuinamente *heavyweight*. Isto é, "*existe um objeto real que está a um metro e meio de distância de mim*" não pode ser caracterizada *heavyweight*. Os mundos mais próximos em que a respectiva proposição é falsa são mundos nos quais não existe qualquer biscoito diante de você – ou mundos nos quais o referido objeto está a mais de uma metro e meio de distância, ou até mesmo mundos nos quais o mesmo objeto foi movido para outro lugar. Nestes mundos particulares, em que o objeto em questão está ausente ou distante, você não irá acreditar existir um biscoito a um metro e meio de distância de você. A conclusão é, portanto, a seguinte: você pode ter razões conclusivas para a proposição em questão, mas a mesma não pode ser caracterizada como *heavyweight*.

Para este caso particular, uma proposição genuinamente *heavyweight* seria, por exemplo, algo como: "*existe um objeto real*"[46]. Agora, que exemplo de alternativa relevante poderíamos fornecer para esta proposição? McBride observa que a alternativa relevante para esta proposição deve ser um mundo no qual não exista um objeto real. Ou seja: um mundo muito estranho, no qual tudo não passa de criação de sua mente que, apesar de muito poderosa, está bastante confusa. Em tais circunstâncias, você acreditaria na proposição "*existe um objeto real*" – que foi baseada em sua "percepção" de um biscoito – mesmo que não existisse tal objeto. Porém, novamente, isso significa dizer que você carece de razões conclusivas para esta proposição.

1.7.1 O desafio não solucionado de Dretske

McBride não alcançou uma solução definitiva para o problema do fecho epistêmico. Ele reconhece isso. Porém, ele realizou algo de grande importância, pois sua resposta a Hawthorne nos mostrou algo curioso sobre o desafio de Dretske: apesar das muitas tentativas de solução, ele ainda continua de pé.

Não podemos nos esquecer do que está em disputa aqui: a aceitação ou rejeição do fecho epistêmico. Não estamos tentando estabelecer uma concepção particular de conhecimento. Assim sendo, ressaltamos que a plausibilidade da teoria das razões conclusivas não é o que está em jogo. Os casos apresentados por Hawthorne tomam a respectiva teoria como o centro das atenções, mas tal decisão não é obrigatória. Aceitamos que a rejeição do fecho por parte de Dretske seja uma consequência necessária de sua própria concepção sobre o conhecimento, que por sua vez se baseia em sua teoria de razões conclusivas. Mas, vale ressaltar, essa não a única razão para rejeitar o fecho. Deste modo, mesmo considerando que a teoria das razões conclusivas seja falsa, o caminho para a rejeição do fecho epistêmico permanece aberto.

A motivação contra o fecho epistêmico pode, na verdade, ser bastante simples: podemos rejeitá-lo simplesmente porque o tomamos por inválido (e não neces-

[46] A proposição original, em inglês, é "*there is a mind-independent object*."

sariamente porque mantemos uma certa concepção de conhecimento que implique em sua rejeição). No caso de Dretske, temos pelo menos duas motivações:

1. A abordagem das razões conclusivas nos compromete com a rejeição do fecho.
2. Não podemos conhecer proposições *heavyweight*.

A questão que se nos apresenta, então, é a seguinte: Para rejeitar o fecho epistêmico, precisamos realmente tomar como base somente essas motivações?

Ora, poderíamos sustentar que não. De qualquer modo, isso não iria provar nada sobre as concepções de Dretske. Ao apresentar ou escolher outras razões, nós não neutralizamos as alternativas restantes. Assim, mesmo que eu, por alguma razão em particular, decida rejeitar o fecho epistêmico por outros meios, eu não necessariamente retiro Dretske do jogo. Ao que parece, o problema do fecho parece seguir do seguinte modo:

Para os defensores do fecho:

Suponha que você seja um defensor do fecho. Então, dado que Dretske é um dos que o rejeitam, você deve encontrar argumentos contra sua tese de que o fecho seja inválido. Considerando que a tese deste último é baseada em 'razões conclusivas' e 'proposições *heavyweight*', você deve mostrar que a respectiva tese é falsa. Porém, como vimos, isso ainda não foi feito – pelo menos de modo totalmente satisfatório.

Para os que rejeitam o fecho:

Suponha que você toma o fecho como inválido. No entanto, você também pensa que as teorias de Dretske são muito problemáticas. Assim, você prefere basear suas ideias em outras razões que não as de Dretske (DE ALMEIDA, 2007). Apesar de tudo, você – nem ninguém até agora – conseguiu encontrar um contraexemplo efetivo para sua tese. Isto é, os contraexemplos fornecidos até agora não são suficientes. Logo, a rejeição do fecho à moda de Dretske ainda está disponível.

Assim, observando o atual estado de coisas, somos forçados a reconhecer que a discussão sobre o fecho, em âmbito epistemológico, se encontra em aberto. Uma das razões para isso, ao que parece, é que praticamente todos os epistemólogos que se debruçam sobre essa questão parecem aceitar muito tacitamente a dicotomia "validade/invalidade" dos princípios de fecho, sem antes questionarem os pressupostos lógicos envolvidos, bem como suas pretensões de aplicação para os respectivos princípios. No próximo capítulo, investigaremos os desdobramentos dessa problemática do fecho na perspectiva da lógica epistêmica, que já se encontra bastante desenvolvida e lida com bastante naturalidade com os princípios de fecho.

A ideia, portanto, é aproveitar estratégias da epistemologia formal (conjunto de teorias formais e lógicas epistêmicas) para trabalhar, em âmbito epistemológico, o problema dos princípios de fecho. No terceiro capítulo, faremos a distinção entre "incognoscibilidade necessária" e "incognoscibilidade contingente". Daí, com base em tal distinção, analisaremos alguns princípios de fecho epistêmico a partir da perspectiva das proposições "contingentemente incognoscíveis".

2 Onisciência lógica: uma perspectiva lógica do problema do fecho epistêmico

"Uma pessoa que conhece qualquer coisa, por esse próprio fato conhece que ela o conhece, e conhece que conhece que o conhece, e assim por diante ad infinitum."

(Baruch Spinoza)

2.1 Introdução

Neste capítulo, o fecho epistêmico será investigado a partir de uma perspectiva formal, de modo a evidenciar aquilo que, em lógica epistêmica, se conhece como "problema da onisciência lógica". A este problema, como veremos, são oferecidas várias soluções. Cada uma delas tem uma finalidade específica na abordagem da questão da onisciência lógica. Devido a este fato, ser-nos-á possível demonstrar que, em lógica epistêmica – ou, como prefere Hendricks (2006), em "epistemologia formal" – o problema gerado pelo fecho é solucionado com base na aplicação da lógica de interesse. Isto é, cada lógica pode invalidar uma quantidade variável de princípios de fecho, segundo sua intenção específica de aplicação.

Isto nos leva, então, ao seguinte resultado: falar em validade ou invalidade (*simpliciter*) do fecho é um erro (e inclusive muito comum) da epistemologia *mainstream*; existem várias formas de fecho e várias formas de onisciência lógica. Cada lógica endossa ou rejeita um determinado tipo de fecho, segundo a aplicação pretendida. Deste modo, os conceitos 'validade' ou 'invalidade' estão sempre restritos ao escopo de uma lógica específica e, portanto, não podem ser extrapolados para além dos seus propósitos. Logo, qualquer teoria sobre princípios de fecho possui como pressuposto inicial (muitas vezes implícito, no caso da epistemologia informal) uma lógica que tenta modelar as capacidades epistêmicas de seus agentes segundo interesses previamente definidos. Minha sugestão, por fim, é a de que devemos fazer a mesma coisa em epistemologia *mainstream* – isto é, restringir a aplicação dos conceitos de validade ou invalidade à lógica utilizada – pois mostro que a questão do fecho e do ceticismo funciona exatamente como uma aplicação de um tipo específico de fecho epistêmico a um contexto particular – o do ceticismo filosófico. Este é o principal objetivo deste capítulo.

Nesta aplicação – isto é, a agentes conjecturadores de hipóteses céticas – sugiro a não-aceitação do fecho. Farei a demonstração disto no capítulo 3, quando já houver demonstrado, por sua vez, a caracterização das hipóteses céticas como "proposições contingentemente incognoscíveis". Por ora, retomarei a problemática do fecho a partir de uma perspectiva formal: a onisciência lógica.

A propriedade de onisciência lógica, como veremos, é uma consequência direta da relação entre duas lógicas, a saber, a lógica modal alética clássica e a lógica epistêmica clássica.

A lógica modal alética – ou simplesmente "lógica modal" – ocupa-se das noções de necessidade e possibilidade lógica; analisa proposições modais que possuem a forma "*é necessário que...*" e "*é possível que...*". Os diferentes sistemas de lógica são caracterizados pelos seus diferentes conjuntos de axiomas. Para fazermos adequadamente a discussão sobre lógica epistêmica, apresentamos alguns dos sistemas mais conhecidos da lógica modal alética, através dos axiomas que os caracterizam. Em lógica modal alética, valem todos os teoremas da lógica clássica; além dos axiomas e regras da lógica proposicional, os sistemas aléticos possuem seus axiomas modais característicos. Tomando o símbolo $\Box$ para representar o operador "*é necessário que...*", temos:

Sistema	**Axioma**
K	$\Box(p \to q) \to (\Box p \to \Box q)$
D	$\Box p \to \neg\Box\neg p$
T	$\Box p \to p$
S4	$\Box p \to \Box\Box p$
S5	$\neg\Box p \to \Box\neg\Box p$

Além desses axiomas, vale também a seguinte regra:

Necessitação: Se p é um teorema, então $\Box p$ também é um teorema.

Semânticas para os respectivos sistemas são apresentadas em livros clássicos de lógica modal, e todas têm, por base, a semântica de mundos possíveis desenvolvida por Kripke[1].

Inspirado no desenvolvimento da lógica modal alética, Hintikka propõe uma abordagem formal para o conhecimento e a crença. Essa abordagem veio a ser conhecida como "lógica epistêmica", chamada assim por lidar com modalidades epistêmicas. Desde seu trabalho pioneiro sobre a lógica epistêmica, *Knowledge and Belief* (publicado em 1962), Hintikka deu início a uma vasta discussão não somente no campo da lógica ou da epistemologia, mas também no campo da inteligência artificial.

Inicialmente, a lógica epistêmica consistiu apenas de uma interpretação diferente dada ao operadores modais já trabalhados em lógica modal alética. O "*é necessário que...*" e o "*é possível que...*", isto é, as modalidades aléticas, são substituídas por Hintikka, que passou a estudar modalidades epistêmicas como "*conhece que...*" e "*acredita que...*". Em 1951, Von Wright já havia iniciado o estudo de modalidades epitêmicas no livro *An Essay in Modal Logic*; mas foi com Hintikka, em 1962, que a lógica epistêmica foi estudada de modo sistemático, e de uma maneira nova até então.

Num âmbito sintático, o sistema epistêmico de Hintikka consiste apenas no aumento da linguagem do cálculo proposicional, a partir dos operadores unários K (para o conhecimento) e B (para a crença). Sendo assim, o conhecimento e a crença são entendidos nessa abordagem como operadores modais (sendo uma espécie de versão epistêmica do operador de necessidade).

No sistema de Hintikka, a expressão "Kp" deve ser compreendida como "é conhecido/sabido que p". O estudo posterior permitiu, inclusive, que fosse possível identificar o agente a quem o conhecimento de p está sendo atribuído. Assim, a expressão "$K_a p$" deve ser compreendida como "o agente a conhece/sabe que p".

A semântica do sistema é baseada naquela já existente para o operador $\Box$. Conhecimento é aquilo que é verdadeiro em todos os mundos *epistemicamente possíveis*. Stalnaker (2006, p. 171) escreve:

> A assunção é a de que ter conhecimento é ter uma capacidade de localizar o mundo real no espaço lógico, para excluir certas possibilidades das candidatas à realidade. As possibilidades epistêmicas

[1] Ver, por exemplo, Chellas (1995), ou Hughes & Cresswell (1996).

> são aquelas que permanecem depois da exclusão, aquelas às quais o conhecedor não consegue distinguir da realidade[2].

Tal como em lógica modal, as possibilidades epistêmicas são definidas por uma relação de acessibilidade entre os mundos possíveis; essa relação é um componente primitivo de um modelo epistêmico. No sistema proposto em *Knowledge and Belief*, temos o seguintes casos:

1. Para o operador K, a relação de acessibilidade entre os mundos é reflexiva e transitiva;

2. Para o operador B, a relação de acessibilidade entre os mundos é reflexiva e serial.

Por ser baseado na lógica modal alética, o sistema de Hintikka herda todas as propriedades dos sistemas correlatos. Por exemplo, os seguintes teoremas da lógica modal alética são também teoremas da lógica epistêmica (com a ressalva de substituição adequada dos símbolos):

Teorema 2.1. *Se a relação R de acessibilidade é serial, então o esquema $\Box p \to \neg\Box\neg p$ é válido.*

Teorema 2.2. *Se a relação R de acessibilidade é reflexiva, então o esquema $\Box p \to p$ é válido.*

Teorema 2.3. *Se a relação R de acessibilidade é transitiva, então o esquema $\Box p \to \Box\Box p$ é válido.*

Há também outros teoremas importantes, um deles é justamente o princípio de fecho (E-CLOS $1'$), que afirma a distribuição do operador $\Box$ sobre a implicação:

Teorema 2.4. *O esquema $\Box(p \to q) \to (\Box p \to \Box q)$ é válido em todos os sistemas de lógica modal alética.*

No sistema proposto em *Knowledge and Belief*, com exceção do Teorema 2.2, todos os teoremas valem para o operador B. Já para o operador K, valem todos os teoremas acima.

[2]*"The assumption is that to have knowledge is to have a capacity to locate the actual world in logical space, to exclude certain possibilities from the candidates for actuality. The epistemic possibilities are those that remain after de exclusion, those that the knower cannot distinguish from actuality."*

Com base na relação de acessibilidade do modelo epistêmico, e nos teoremas apresentados acima, já podemos perceber que todos os axiomas abaixo são aceitos pelo sistema de Hintikka:

1. $K_a(p \to q) \to (K_a p \to K_a q)$;
2. $K_a p \to \neg K_a \neg p$;
3. $K_a p \to p$;
4. $K_a p \to K_a K_a p$;
5. $B_a(p \to q) \to (B_a p \to B_a q)$;
6. $B_a p \to \neg B_a \neg p$;
7. $B_a p \to B_a B_a p$.

Cada um desses esquemas carrega um certo comprometimento em relação ao modo como o conhecimento e a crença são entendidos. Por exemplo, o esquema $K_a p \to K_a K_a p$ afirma que o agente a possui capacidades introspectivas. Isto é, se o agente a conhece uma proposição p, então o agente a sabe que sabe disso; isto é, o agente sempre terá conhecimento sobre sua base de conhecimento. Esta é uma propriedade bastante discutida na epistemologia; alguns contraexemplos são propostos na literatura sobre o tema, sugerindo, pelo menos, uma limitação na aplicabilidade de $K_a p \to K_a K_a p$. Apesar disso, em interpretações bem particulares do operador de conhecimento, $K_a p \to K_a K_a p$ mostra-se aceitável[3]. A versão doxástica do Axioma **S4**, isto é, o axioma $B_a p \to B_a B_a p$, comporta questões semelhantes.

Vale ressaltar que a questão de saber se o Axioma **S4** é aceitável (seja ele para o conhecimento ou para a crença) já leva a uma outra questão, que é a de definir que tipo de conhecimento se está a falar quando se considera a introspecção; isto é, que tipo de conhecimento é compatível com o princípio lógico evidenciado pelo Axioma **S4**. Claramente, o Axioma **S4** não é apropriado para modelar o conhecimento do senso comum: o famoso personagem Forrest Gump não tinha conhecimento de que sabia jogar *ping-pong* tão bem, apesar de ser um profissional no jogo. De modo similar, uma pessoa que tem certos preconceitos pode acreditar que alguns indivíduos são inferiores a outros, apesar de não acreditar que tem essas

[3]Ver MALCOLM, Norman. Knowledge and belief. **Mind** (new series). Vol. 61, n. 242, p. 178-189, abr. 1951.

crenças; é o caso do preconceito implícito: o sujeito possui crenças preconceituosas mas, ao mesmo tempo, não acredita que as possui.

O axioma $K_a p \to p$ é amplamente aceito: aquilo que é conhecido deve ser verdadeiro. Apesar de haver muita discussão sobre quais critérios são necessários para se definir conhecimento, uma coisa é certa: a verdade é um dos seus componentes essenciais. Deste modo, não pode haver algo que seja conhecido e seja ao mesmo tempo falso. Já a versão doxástica do mesmo axioma, isto é, o axioma $B_a p \to p$, é claramente inaceitável. Ora, nem sempre tudo que acreditamos é verdadeiro: um agente pode ter uma crença formulada em p, acreditando que p é verdadeira, sendo p, ao invés disso, uma proposição claramente falsa.

Assim como o axioma $K_a p \to p$, os axiomas $K_a p \to \neg K_a \neg p$ e $B_a p \to \neg B_a \neg p$ não geram muita discussão. Ora, se um agente conhece que p é o caso, ele não pode conhecer que p não seja o caso; isto é, se o agente soubesse que $\neg p$ fosse o caso, então $\neg p$ seria verdadeiro, mas isso geraria uma contradição: pois a hipótese inicial é de que o agente conhece p, e que portanto p deve ser verdadeiro. Como a lógica epistêmica de Hintikka é apenas uma extensão da clássica, ela mantém o princípio da não-contradição; logo, duas sentenças contraditórias não podem ser verdadeiras ao mesmo tempo. Algo similar se aplica à crença.

Os dois axiomas restantes, a saber, $K_a(p \to q) \to (K_a p \to K_a q)$ e $B_a(p \to q) \to (B_a p \to B_a q)$, são, como pudemos constatar no primeiro capítulo, alvos de grande discussão epistemológica. Acontece, no entanto, que os problemas gerados por estes princípios de fecho não estão restritos apenas ao âmbito informal da epistemologia "*mainstream*". Estes dois esquemas geram aquilo que, em lógica epistêmica (ou epistemologia formal), ficou conhecido como "o problema da onisciência lógica". Este problema, objeto de estudo deste capítulo, será agora investigado, como dissemos, a partir de uma perspectiva formal.

Como podemos notar, os axiomas $K_a(p \to q) \to (K_a p \to K_a q)$ e $B_a(p \to q) \to (B_a p \to B_a q)$ são a versão epistêmica ou doxástica do conhecido axioma modal K, $\Box(p \to q) \to (\Box p \to \Box q)$. O axioma K é o mais fraco da lógica modal alética, característico do sistema K. Os sistemas de lógica modal T, $S4$, $S5$ são todos extensões do sistema K, através da adição dos axiomas característicos T, $S4$ e $S5$. Diversos outros sistemas são obtidos pela combinação ou adição desses axiomas ou de outros. O importante a se notar aqui é que o axioma K é válido em todos esses sistemas: ele é obtido pelo modo característico de interpretar o operador modal $\Box$ e o operador clássico $\rightarrow$.

O sistema de Hintikka possui as mesmas características de um sistema padrão de lógica modal; a diferença é que, com ele, são estudadas as noções de conhecimento e crença, ao invés da noção de necessidade lógica. Pelas características semelhantes aos sistemas de lógica modal alética, o sistema de Hintikka é levado a aceitar de início axiomas como $K_a(p \rightarrow q) \rightarrow (K_a p \rightarrow K_a q)$ e $B_a(p \rightarrow q) \rightarrow (B_a p \rightarrow B_a q)$ – as versões epistêmica e doxástica do princípio (E-CLOS $1'$). No entanto, como pudemos observar no capítulo 1, esse princípio é, epistemologicamente falando, muito forte para agentes epistêmicos com capacidades epistêmicas reais – isto é, limitadas.

O que dizer de um ser humano com capacidades racionais perfeitas? Isto é, o que dizer de um indivíduo capaz de deduzir todas as consequências lógicas do seu conhecimento atual? Certamente, hoje em dia, esse tipo de pessoa seria considerada algo mais do que um simples ser humano. Mesmo os maiores gênios consagrados pela história não tiveram capacidade de saber, em sua época, todas as possibilidades de aplicação de suas invenções. Como exemplo, temos Leonardo da Vinci. Planejando repelir os navios turcos de Veneza, ele esboçou a confecção de roupas de mergulho para os soldados italianos, para que pudessem, por debaixo d'água, furar os cascos dos navios inimigos, afundando-os. A roupa primitiva de mergulho nunca foi de fato confeccionada por Da Vinci. Porém, mais de trezentos anos depois, o *Discovery Channel* mostrou que a roupa de fato funciona debaixo d'água. Talvez não teria sido muito útil como equipamento militar na época, mas Da Vinci mal sabia que sua invenção daria um excelente equipamento de exploração das águas de Veneza (ou dos oceanos, de um modo geral). Apesar de ser uma consequência lógica do conhecimento de Da Vinci, a ideia da roupa de mergulho para fins de exploração não chegou a ser conhecida dele.

Isso mostra, portanto, que mesmo seres humanos que gozam de um nível considerável de capacidade criativa não são capazes de inferir todas as consequências lógicas de seu conhecimento. O mesmo ocorre com as máquinas "inteligentes"; ainda que possuam bons recursos computacionais, elas não são capazes de computar todas as consequências lógicas de seus bancos de dados.

A questão é: como uma teoria lógica deve proceder com relação ao problema da onisciência lógica? Essa pergunta nos leva a vários pontos de investigação. Primeiramente, devemos nos perguntar a quem o problema da onisciência lógica deve ser endereçado. Isto é, o problema da onisciência lógica é um problema para quem? A teoria epistêmica de Hintikka é adequada para os agentes humanos (ou artificiais) com capacidades racionais limitadas? Como devemos entender uma teoria lógica para o conhecimento ou para a crença? Uma teoria lógica epistêmica deve consistir de um modelo lógico capaz de esclarecer o que seja a crença e o conhecimento, representando fielmente as capacidades racionais reais dos agentes? Ou devemos entender uma teoria lógica epistêmica como um modelo lógico para uma teoria epistemológica qualquer? Tanto os seres humanos quanto as máquinas falham em ser logicamente oniscientes por vários motivos[4]. Segue-se daí outra questão: é necessário, para uma lógica epistêmica, satisfazer todos os motivos de falha de onisciência lógica? Ou será que essa lógica pode ser uma idealização, satisfazendo um motivo particular qualquer? Se isso for o caso, há certas ocasiões nas quais a propriedade de onisciência lógica é aceitável. Por último, fazemos a seguinte questão: é a onisciência lógica de fato um problema, ou isto irá depender das pretensões de modelagem da lógica epistêmica que lidará com essa propriedade?

Neste capítulo, tentarei cobrir estas questões. Para isto, serão apresentadas algumas soluções encontradas na literatura da lógica epistêmica acerca do problema da onisciência lógica. Entretanto, como sabemos, a discussão sobre este problema é muito ampla; há um número considerável de soluções para o problema da onisciência lógica, tão grande que, assim reconheço, a tarefa de analisá-las todas é praticamente impossível. Apesar disto, os esforços serão concentrados nas principais abordagens, aquelas que conquistaram espaço e que, como se verificará, possuem grande peso na literatura sobre o assunto.

Este capítulo será divido em cinco sessões, sendo uma introdutória, três principais e uma conclusiva. Estas estão organizadas do seguinte modo:

1. Sessão introdutória (esta sessão), que introduz o tema e o objetivo do capítulo, bem como algumas noções relevantes para discussão do problema da onisciência lógica.

2. A segunda sessão é centrada na abordagem original da lógica epistêmica, tendo como foco principal o próprio Hintikka. Nesta sessão, demonstramos três casos de onisciência lógica no sistema de Hintikka, além de algumas propriedades adicionais. A solução de Hintikka para o problema da onisciência,

[4]Alguns desses motivos já foram expostos no capítulo 1.

proposta em 1975 no artigo *Impossible possible worlds vindicated*, também é examinada. Aqui, demonstramos também propriedades relativas a essa abordagem, com base em uma semântica padrão de mundos possíveis. Tanto a abordagem original, quanto a dos "impossíveis mundos possíveis" de Hintikka lidam com o problema da onisciência lógica por meio do apelo semântico; por esta razão, resolvemos deixá-las juntas nesta sessão.

3. A terceira abordagem, a saber, a lógica das crenças implícitas e explícitas, é estudada na terceira sessão, juntamente com outras duas: lógica da consciência e lógica da consciência geral. A terceira sessão segue apresentando essas lógicas, mostrando como elas lidam com o problema da onisciência lógica. Também nela demonstramos as propriedades referentes à onisciência lógica. Essas lógicas são organizadas em sessões próximas pelo seguinte motivo: todas elas fazem a distinção entre "crenças implícitas" e "crenças explícitas" (distinção que não aparece em Hintikka). A lógica das crenças implícitas e explícitas também lida com o problema da onisciência lógica por meio de um apelo semântico; por isso, resolvemos deixá-la mais próxima às abordagens anteriores de Hintikka. As outras duas, isto é, a lógica da consciência e a lógica da consciência geral, já possuem uma outra forma de lidar com o problema, combinando elementos sintáticos com elementos semânticos. Resolvemos, por conta disto, deixá-las mais próximas do capítulo seguinte, que trata das abordagens puramente sentenciais para o problema da onisciência lógica.

4. A quarta sessão deste capítulo tem o foco central no modelo dedutivo de crenças apresentado por Konolige em 1984, em sua tese de doutorado cujo título é "*A deduction model of belief and its logics*". Aqui, apresentamos o modelo dedutivo de crenças, e explicitamos algumas propriedades apresentadas no trabalho de Konolige.

5. Por fim, na quinta sessão, fazemos uma síntese sobre tudo aquilo que foi apresentado no decorrer do capítulo. Aqui, sistematizamos as perguntas importantes a que temos intenção de responder, e oferecemos suas respectivas respostas após os argumentos. Esta é a última parte do capítulo. Espera-se que, neste ponto, a necessidade de se abordar o problema do fecho epistêmico a partir da perspectiva de "aplicação da lógica subjacente" tenha sido demonstrada.

2.2 A lógica epistêmica de Hintikka: o problema da onisciência lógica

O problema da onisciência lógica tem sido muito discutido desde as primeiras formulações da lógica epistêmica. Em *Knowledge and Belief* (1962), Hintikka propõe uma abordagem para a lógica epistêmica na qual, como é possível verificar, a fórmula $K_a p \to K_a q$ é válida tão logo a implicação lógica de p para q seja válida proposicionalmente. Assim, supondo $\vDash p \to q$ e $\vDash K_a p$, temos $\vDash K_a q$[5].

No decorrer deste capítulo, as regras de Hintikka serão estudadas em sua forma original. Após a exposição das regras, demonstraremos consequências importantes que se seguem a partir delas. Entre as consequências, estão esquemas que atribuem aos agentes a propriedade de onisciência lógica.

A solução ao problema da onisciência lógica, apresentada na semântica dos "impossíveis mundos possíveis" (HINTIKKA, 1975), também será discutida. Utilizaremos uma semântica padrão de mundos possíveis, baseada nas estruturas de Kripke, para demonstrar propriedades e discutir alguns resultados desta abordagem[6].

2.2.1 Os diferentes casos de onisciência lógica

Para os propósitos deste capítulo, daremos atenção especial a estes quatro casos específicos de onisciência lógica – já discutidos no capítulo 1:

1. **Onisciência lógica total (E-CLOS 2)**. Um agente é total-logicamente onisciente com relação a uma classe **E** de estruturas se, sempre que ele conhece todas as fórmulas de um conjunto Γ ; e Γ implica logicamente a fórmula φ com relação a **E**, então o agente também conhece φ.

2. **Fecho sob implicação material (E-CLOS 1)**. Quando $K_a q$ é obtida a partir de $K_a p$ e $K_a(p \to q)$. Ou seja, se o agente a conhece p e conhece $p \to q$, então o agente a também conhece q. O fecho sob implicação material é expresso,

[5] A notação utilizada não é a mesma do *Knowledge and Belief* por motivo de simplificação, já que analisaremos várias abordagens diferentes acerca do mesmo tema.

[6] As definições utilizadas nessa semântica são adaptações das definições apresentadas por FITTING, Melvin; MENDELSOHN, Richard L. **First-order modal logic**. Dordrecht: Kluwer Academic Publishers, 1998.

portanto, em esquemas como $K_a(p \to q) \to (K_a p \to K_a q)$, $(K_a p \wedge K_a(p \to q)) \to K_a q$, entre outros.

3. **Fecho sob implicação válida (E-CLOS 6)**. Quando $K_a q$ é obtida a partir de $K_a p$ e da implicação $p \to q$. Isto é, se $p \to q$ é uma fórmula proposicionalmente válida, e o agente a conhece p, então o agente a também conhece q. Apesar de não ser exatamente a mesma coisa, o fecho sob implicação válida é expresso em esquemas como $(K_a p \wedge p \to q) \to K_a q$. Dizemos não ser exatamente a mesma coisa porque o referido esquema não especifica que a implicação $p \to q$ vale em todas as circunstâncias possíveis – ou seja, que é uma fórmula válida. Deste modo, se quisermos enunciar o fecho sob implicação válida de maneira precisa, usamos: "se na circunstância atual c, $c \models K_a p$ e $\models p \to q$ – ou seja, é válida – então $c \models K_a q$.

4. **Conhecimento de fórmulas válidas (E-CLOS 3)**. $K_a p$ é obtida a partir de p. Ou seja, se p é uma fórmula válida, então o agente a conhece p.

As consequências decorrentes da aceitação destas propriedades são muito sérias; a lógica epistêmica de Hintikka garante aos seus agentes capacidades epistêmicas infinitas. Para tornar as coisas mais claras, vejamos o seguinte exemplo de Fitting & Mendelsohn (1998, p. 29):

> Suponha, por exemplo, que a Conjectura de Goldbach seja provável (apesar de que ainda não tenhamos qualquer prova). Então, qualquer um que saiba que os axiomas de Peano são verdadeiros – provavelmente, qualquer leitor deste livro – sabe, por onisciência lógica, que a Conjectura de Goldbach é verdadeira. Ainda que seja improvável que qualquer leitor deste livro esteja de posse de uma prova da Conjectura de Goldbach, também é improvável que qualquer leitor deste livro saiba que a conjectura de Goldbach é verdadeira, e mais improvável ainda que qualquer leitor deste livro saiba que ele sabe que a conjectura de Goldbach seja verdadeira[7].

O exemplo é bem claro. Suponha que os axiomas de Peano sejam verdadeiros. Suponha também que a Conjectura de Goldbach seja verdadeira. Ora, os examinadores deste trabalho de tese conhecem os axiomas de Peano. Logo, os

[7]*"Suppose, for example, that Goldbach's Conjecture is provable (although we have no proof as yet). Then, anyone who knows that Peano Axioms are true – most likely, any reader of this book knows, by logical omniscience, that Goldbach's Conjecture is true. Yet it is unlikely that any reader of this book is in possession of a proof of Goldbach's Conjecture, so it is unlikely that any reader of this book knows that Goldbach's Conjecture is true, and even more unlikely that any reader of this book knows that he knows that Goldbach's Conjecture is true."*

examinadores deste trabalho de tese sabem que a Conjectura de Goldbach é verdadeira. Vejamos ainda uma particularidade mais séria, o caso $\Gamma = \emptyset$ do princípio (E-CLOS 2). Ora, se o agente a conhece todas as fórmulas do conjunto Γ, φ segue logicamente de Γ e $\Gamma = \emptyset$, então a conhece todas as consequências de $\emptyset$; isto é, a conhece todos os teoremas. Certamente, estes dois exemplos demonstram que a lógica epistêmica de Hintikka se compromete com teoremas muito fortes.

Demonstraremos, no decorrer deste capítulo (seção 2.2.3), que os fechos sob implicação material, implicação válida, conhecimento de fórmulas válidas e onisciência lógica total são todos deriváveis no sistema de Hintikka.

2.2.2 As regras de Hintikka em *Knowledge and Belief*

Veremos a seguir, de um modo simplificado, as regras em sua forma original. Com elas, demonstraremos algumas propriedades relativas ao operador K, incluindo três dos quatro casos apresentados de onisciência lógica – a saber, os princípios (E-CLOS 1), (E-CLOS 3) e (E-CLOS 6). Entre as propriedades demonstradas, estarão também esquemas bastante discutidos não somente na lógica, como também na filosofia e epistemologia, é o caso da introspecção positiva ($K_a p \rightarrow K_a K_a p$)[8].

As regras de Hintikka podem ser entendidas como condições de consistência ou, melhor ainda, como condições de "defensilibidade"[9]. Um conjunto λ de sentenças é indefensável se, e somente se, não pode ser subconjunto de um conjunto μ de sentenças que satisfaz as seguintes condições:

(Proposicionais)

1. $(C.\neg)$ Se $p \in \mu$, então $\neg p \notin \mu$.

2. $(C.\wedge)$ Se $(p \wedge q) \in \mu$, então $p \in \mu$ e $q \in \mu$.

[8] A introspecção negativa, identificada pelo esquema $\neg K_a p \rightarrow K_a \neg K_a p$, também é possível na lógica de Hintikka. Para isso, é necessário adicionar uma regra estabelecendo que a relação entre conjuntos de sentenças seja também euclideana. Vale ressaltar, contudo, que este é um princípio rejeitado por ele. A relação entre esses conjuntos de sentenças, por exemplo, μ e μ^*, é definida por Hintikka como "relação de alternância". Em lógica modal alética, é comum que ela seja chamada de "relação de acessibilidade" (HINTIKKA, 1962, p. 106).

[9] *Defensibility* termo utilizado por Hintikka para evitar erros de interpretação contra as (A) *rules* (as regras propriamente ditas), estabelecidas no capítulo 2 do *Knowledge and Belief*. As (C) *rules*, que são as regras que mostraremos, são as (A) *rules* apresentadas, de maneira mais sistemática, no capítulo 3 dessa mesma obra.

3. $(C.\vee)$ Se $(p \vee q) \in \mu$, então $p \in \mu$ ou $q \in \mu$.

4. $(C.\neg\neg)$ Se $\neg\neg p \in \mu$, então $p \in \mu$.

5. $(C.\neg\wedge)$ Se $\neg(p \wedge q) \in \mu$, então $\neg p \in \mu$ ou $\neg q \in \mu$.

6. $(C.\neg\vee)$ Se $\neg(p \vee q) \in \mu$, então $\neg p \in \mu$ e $\neg q \in \mu$.

(Epistêmicas)

Um sistema modelo Ω é um conjunto de conjuntos de sentenças μ que satisfazem os requisitos proposicionais acima. Seja $P_a =_{def} \neg K_a \neg$. Neste caso, valem as seguintes propriedades:

7. $(C.P^*)$ Se $P_a p \in \mu$ e se μ pertence a um sistema modelo Ω, então há em Ω ao menos uma alternativa μ^* a μ (com relação a a) tal que $p \in \mu^*$.

8. $(C.KK^*)$ Se $K_a q \in \mu$ e se μ^* é qualquer alternativa a μ (com relação a a) em um sistema modelo Ω, então $K_a q \in \mu^*$.

9. $(C.K)$ Se $K_a p \in \mu$, então $p \in \mu$.

10. $(C.\neg K)$ Se $\neg K_a p \in \mu$, então $P_a \neg p \in \mu$.

11. $(C.\neg P)$ Se $\neg P_a p \in \mu$, então $K_a \neg p \in \mu$.

Acrescentaremos mais duas regras proposicionais. Essas regras facilitam e encurtam as provas[10]:

12. (C. $\rightarrow$) Se $p \rightarrow q \in \mu$, então, se $p \in \mu$ então $q \in \mu$.

13. $(C.\neg \rightarrow)$ Se $\neg(p \rightarrow q) \in \mu$, então $p \in \mu$ e $\neg q \in \mu$.

Há uma série de outras regras derivadas destas primeiras. São chamadas

[10]Apesar de não estarem entre as originais, estas regras não alteram o sistema; para provar isto, basta observar que são regras clássicas, e que o sistema de Hintikka é uma expansão da lógica clássica.

de *condições alternativas* (HINTIKKA, 1962, p. 44). O uso dessas regras varia de acordo com a preferência de cada um (o importante é que elas têm exatamente o mesmo efeito que as outras). Se aceitarmos as condições $(C.KK^*)$ e $(C.K)$, podemos derivar a partir delas a seguinte regra adicional:

$(C.K^*)$ Se $K_a p \in \mu$ e μ^* é uma alternativa a μ (com relação a a), em algum sistema modelo, então $p \in \mu^*$.

Hintikka diz que, "se $(C.K^*)$ for utilizada, $(C.K)$ pode obviamente ser substituída pela seguinte condição..." (HINTIKKA, 1962, p. 45):

$(C.refl)$ A relação de alternância é reflexiva.

Já é conhecido, em lógica modal, que se a relação de alternância é reflexiva, então ela também é serial. Daí, a partir de $(C.refl)$, derivamos a regra:

$(C.min)$ Em todo sistema modelo, cada conjunto modelo tem ao menos uma alternativa.

Intuitivamente, a regra $(C.min)$ quer dizer que, em todos os mundos possíveis (circunstâncias epistêmicas atuais), o agente a sempre considera algum mundo possível (circunstância epistêmica alternativa à atual). Há também, a partir de $(C.K^*)$ e $(C.min)$, a seguinte condição:

$(C.k^*)$ Se $K_a p \in \mu$ e se μ pertence a um sistema modelo Ω, então há em Ω ao menos uma alternativa μ^* a μ (com relação a a) tal que $p \in \mu^*$.

Além das condições até então apresentadas, podemos ter ainda outra condição adicional, que na verdade é uma formulação diferente da condição $(C.KK^*)$:

$(C.trans)$ Se μ_2 é uma alternativa a μ_1 e μ_3 a μ_2, ambos com relação ao mesmo a, então μ_3 é uma alternativa a μ_1, com relação a a.

A condição $(C.trans)$ estipula que a relação de alternância é transitiva.

2.2.3 Resultados em *Knowledge and Belief*: onisciência lógica e outros

A partir das regras apresentadas, é possível demonstrar uma série de propriedades interessantes:

Teorema 2.5. $\models K_aK_bp \to K_ap$ *(transmissão de conhecimento)*

1. $K_aK_bp \in \mu$	Hipótese
2. $\neg K_ap \in \mu$	Hipótese
3. $P_a\neg p \in \mu$	2, $(C.\neg K)$
4. $\neg p \in \mu^*$	3, $(C.P^*)$
5. $K_bp \in \mu$	1, $(C.K)$
6. $K_bp \in \mu^*$	5, $(C.KK^*)$
7. $p \in \mu^*$	6, $(C.K)$
Contrad.	4,7.

Teorema 2.6. $\models K_ap \to P_ap$ *(o que é conhecido é considerado possível)*

1. $K_ap \in \mu$	Hipótese
2. $\neg P_ap \in \mu$	Hipótese
3. $K_a\neg p \in \mu$	2, $(C.\neg P)$
4. $\neg p \in \mu$	3, $(C.K)$
5. $p \in \mu$	1, $(C.K)$
Contrad.	4,5

Teorema 2.7. $\models K_ap \leftrightarrow K_aK_ap$

$2.3_a \models K_ap \to K_aK_ap$ *(introspecção positiva)*

1. $K_ap \in \mu$	Hipótese
2. $\neg K_aK_ap \in \mu$	Hipótese
3. $P_a\neg K_ap \in \mu$	2, $(C.\neg K)$
4. $\neg K_ap \in \mu^*$	3, $(C.P^*)$
5. $K_ap \in \mu^*$	1, $(C.KK^*)$
Contrad,	4,5.

$2.3_b \models K_aK_ap \to K_ap$ *(simplificação do conhecimento)*

1. $K_a K_a p \in \mu$ Hipótese
2. $\neg K_a p \in \mu$ Hipótese
3. $P_a \neg p \in \mu$ 2, $(C.\neg K)$
4. $\neg p \in \mu^*$ 3, $(C.P^*)$
5. $K_a p \in \mu$ 1,$(C.K)$
6. $K_a p \in \mu^*$ 5, $(C.KK^*)$
7. $p \in \mu^*$ 6, $(C.K)$

Contrad. 4,7.

Teorema 2.8. *A lógica epistêmica de Hintikka contém a lógica clássica.*

Para mostrar isso, basta verificar que, na lógica epistêmica de Hintikka:

- $\models p \to (q \to p)$;
- $\models (p \to (q \to r)) \to ((p \to q) \to (p \to r))$;
- $\models (\neg p \to \neg q) \to ((\neg p \to q) \to p)$;
- Se $\models p$ e $\models p \to q$, então $\models q$.

Deste modo, as regras de Hintikka permitem demonstrar todos os teoremas da lógica clássica.

Vale ressaltar que nem todas as regras apresentadas na seção anterior são necessárias ao sistema de Hintikka. O esquema $K_a(p \to q) \to (K_a p \to K_a q)$ (E-CLOS 1), um dos casos de onisciência lógica, só precisa basicamente de duas regras epistêmicas para ser demonstrado. O mesmo acontece com todos os outros casos; ou melhor, acontece com todos os teoremas da lógica epistêmica padrão de Hintikka.

Com base nas regras apresentadas, já podemos demonstrar três dos quatro casos de onisciência lógica expostos no início desta seção. O princípio de onisciência lógica total, (E-CLOS 2), pode ser demonstrado normalmente em qualquer formulação mais recente da lógica epistêmica de Hintikka, isto é, utilizando uma notação formal mais comum à literatura atual sobre lógica modal. Como estamos trabalhando com as regras originais de *Knowledge and Belief*, deixaremos sua demonstração para outra ocasião. O segundo deles, isto é, o fecho sob implicação material (E-CLOS 1), é percebido e comentado por Hintikka ainda em *Knowledge and Belief*. Os outros dois, a saber, o fecho sob implicação válida e o conhecimento de fórmulas válidas, passam despercebidos por ele, mas têm também como efeito a onisciência lógica dos agentes.

Teorema 2.9. *$\models K_a(p \to q) \to (K_a p \to K_a q)$ (fecho sob implicação material)*

1. $K_a(p \to q) \in \mu$	Hipótese
2. $\neg(K_a p \to K_a q) \in \mu$	Hipótese
3. $K_a p \in \mu$ e $\neg K_a q \in \mu$	2, $(C.\neg \to)$
4. $P_a \neg q \in \mu$	3, $(C.\neg K)$
5. $\neg q \in \mu^*$	4, $(C.P^*)$
6. $K_a(p \to q) \in \mu^*$	1,$(C.KK^*)$
7. $p \to q \in \mu^*$	6, $(C.K)$
8. $K_a p \in \mu^*$	3, $(C.KK^*)$
9. $p \in \mu^*$	8, $(C.K)$
10. $q \in \mu^*$	7,9 $(C.K)$
Contrad.	5, 10

Teorema 2.10. *Se $K_a p \in \mu$ e $\models p \to q$, então $K_a q \in \mu$ (fecho sob implicação válida)*

1. $K_a p \in \mu$	Hipótese
2. $\models p \to q$	Hipótese
3. $\neg K_a q \in \mu$	Hipótese
4. $K_a p \in \mu^*$	$1, (C.KK^*)$
5. $p \in \mu^*$	$4, (C.K)$
6. $P_a \neg q \in \mu$	$3, (C.\neg K)$
7. $\neg q \in \mu^*$	$6, (C.P^*)$
8. $p \to q \in \mu^*$	Hipótese 2
9. $q \in \mu^*$	$5, 8$ $(C. \to)$
Contrad.	7,9

Teorema 2.11. *Se $\models p$, então vale $\models K_a p$ (conhecimento de fórmulas válidas)*

Suponha $\models p$. Deste modo, p vale em todo conjunto modelo μ de todo sistema modelo Ω. Seja μ então um conjunto modelo qualquer de um sistema modelo Ω qualquer. Se p é válida, então $p \in \mu$. Pela regra $(C.min)$, há em Ω ao menos uma alternativa epistêmica μ^* a μ. Novamente, se p é valida, obtemos $p \in \mu^*$. Isso ocorrerá para qualquer μ^*, já que p é uma fórmula válida. Logo, $K_a p \in \mu$.

Com a lógica epistêmica de Hintikka, também é possível demonstrar uma série de outros resultados interessantes. Entre eles, estão[11]:

[11]Os esquemas abaixo são apresentados e discutidos, no âmbito da lógica deôntica, por CRUZ, A. M. P. **Lógica deôntica paraconsistente: paradoxos e dilemas.** Natal(RN): EDUFRN, 2005.

1. $K_a(p \wedge q) \leftrightarrow (K_a p \wedge K_a q)$ *(distribuição de K)*;
2. $P_a(p \vee q) \leftrightarrow (P_a p \vee P_a q)$ *(distribuição de P)*;
3. $(K_a p \vee K_a q) \rightarrow K_a(p \vee q)$;
4. $(P_a p \wedge K_a q) \rightarrow P_a(p \wedge q)$;
5. $p \rightarrow P_a p$;
6. $K_a p \rightarrow P_a p$;
7. $K_a \neg p \rightarrow \neg K_a p$;
8. $P_a(p \wedge \neg p) \rightarrow (p \wedge \neg p)$;
9. $(K_a p \wedge K_a \neg p) \rightarrow q$;
10. $\neg(K_a(p \vee q) \wedge (K_a \neg p \wedge K_a \neg q))$;
11. $\neg p \rightarrow (p \rightarrow K_a q)$;
12. $K_a q \rightarrow (p \rightarrow K_a q)$;
13. $K_a \neg p \rightarrow K_a(p \rightarrow q)$;
14. $K_a q \rightarrow K_a(p \rightarrow q)$.

A lógica epistêmica de Hintikka é interessante não apenas por ser capaz de demonstrar todas as propriedades apresentadas até então. Através de um método proposto por ele mesmo, é possível provar a invalidade de certas fórmulas que representam sentenças auto-contraditórias na linguagem natural (HINTIKKA, 1962, p. 64–76). Um exemplo disso é o conhecido problema de Moore: o de afirmar e desacreditar[12]. Considere então a seguinte sentença:

1. p, mas eu não acredito que p.

Hintikka comenta (1962, p. 64) que a maioria dos primeiros estudiosos do problema de Moore argumentou que a sentença acima não é auto-contraditória. Uma das razões para essa opinião é que, se construída em terceira pessoa, o resultante é uma sentença completamente natural:

[12]Ver MOORE, G. E. **Ethics.** London: Williams & Norgate, 1912. p. 125; A Reply to My Critics. *In*: **The Philosophy of G. E. Moore**, editado por P. A. Schilpp. Evanston: Northwestern University, 1942. p. 541-543; **Philosophical Papers.** Londres: Allen & Unwin, 1959. p. 151-195.

2. p, mas o agente a não agredita que p.

Essa opinião não é aceita por Hintikka. Apesar de reconhecer a naturalidade da sentença 2, ele claramente rejeita 1, colocando-a como auto-contraditória. Ora, se 1 é uma sentença auto-contraditória na linguagem natural, a sua lógica epistêmica também deveria invalidá-la. O modo que Hintikka encontrou para fazê-lo foi utilizando uma paráfrase. De acordo com ele, a proposição expressa por 1 seria a mesma que a proposição expressa por:

3. Eu acredito que o caso é o seguinte: p, mas eu não acredito que p.

Formalizando a sentença, temos $B_a(p \wedge \neg B_a p)$. A partir daí, não é difícil demonstrar sua invalidade[13].

Teorema 2.12. $\not\models B_a(p \wedge \neg B_a p)$

1. $B_a(p \wedge \neg B_a p) \in \mu$	Hipótese
2. $p \wedge \neg B_a p \in \mu$	1, $(C.b^*)$
3. $B_a(p \wedge \neg B_a p) \in \mu^*$	1, $(C.BB^*)$
4. $\neg B_a p \in \mu^*$	2, $(C.\wedge)$
5. $C_a \neg p \in \mu^*$	4, $(C.\neg B)$
6. $\neg p \in \mu^{**}$	5, $(C.C^*)$
7. $p \wedge \neg B_a p \in \mu^{**}$	3, $(C.B^*)$
8. $p \in \mu^{**}$	7, $(C.\wedge)$
Contrad. 6,8	

Este método de Hintikka facilita muito a análise de sentenças "esquisitas" como as que foram mostradas. O interessante é que, enquanto a invalidade de 1 é evidenciada, a naturalidade da sentença 2 também é mantida; ou seja, 2 é uma sentença válida. Para mostrar isso, basta utilizar o mesmo recurso:

4. Eu acredito que o caso é o seguinte: p, mas o agente a não acredita que p.

A paráfrase da sentença 2 resulta na fórmula $B_b(p \wedge \neg B_a p)$. É fácil mostrar $\models B_b(p \wedge \neg B_a p)$.

[13]As regras para a crença (operador B) são basicamente as mesmas para o conhecimento. A única regra não satisfeita no sistema de crenças de Hintikka é a regra $(C.B)$, correspondente da regra $(C.K)$. Sendo assim, nem tudo aquilo que se acredita é verdadeiro. Considere-se também que, quando em relação às crenças, utilizamos o operador C ao invés de P. $C_a =_{def} \neg B_a \neg$.

Já foi dito que não precisamos de todas as regras epistêmicas para derivar os teoremas apresentados. As provas poderiam ser efetuadas se o sistema possuísse, uma por vez, as seguintes combinações de regras:

1. $(C.K)\&(C.KK^*)$;
2. $(C.K)\&(C.K^*)\&(C.trans)$;
3. $(C.refl)\&(C.K^*)\&(C.trans)$;
4. $(C.refl)\&(C.K^*)\&(C.KK^*)$.

Logo, nenhuma dessas formulações escapa ao problema da onisciência lógica. Seja em qualquer combinação das regras acima, os três casos de onisciência lógica surgem como teoremas do sistema. Com isto, fica claro que, com tal abordagem lógica das noções de crença e conhecimento[14], a onisciência lógica passa a ser uma propriedade absolutamente comum a todos os agentes.

2.2.4 Discussão preliminar sobre onisciência lógica

É possível observar que há uma grande discrepância entre as capacidades racionais dos agentes considerados em *Knowledge and Belief* e aquelas dos agentes "reais", ou seja, nós mesmos. De maneira alguma podemos dizer que conhecemos todas as fórmulas válidas da lógica (ou melhor, de todas as lógicas que existem atualmente) ou, então, que conhecemos todas as verdades necessárias do universo – caso exista alguma.

O fascinante desse problema são as várias consequências que traz às várias áreas de conhecimento. Se a onisciência lógica é uma propriedade fora do alcance dos seres humanos e também para a inteligência artificial, então a origem do problema deve rapidamente ser identificada. A identificação daquilo que leva ao problema da onisciência lógica passa então a ser a identificação daquilo que nos faz ser tão limitados racionalmente. O interesse de apenas apontar que a formulação de Hintikka não considerava agentes "reais" levou a uma pesquisa vasta sobre as

[14]As propriedades de onisciência lógica foram demonstradas apenas para a noção de conhecimento. Contudo, como observa Hintikka:
"*We may discuss the notion of belief in much the same way as the notion of knowledge has just been discussed. (...) most of the argument would consist in retracing our steps anyway in a slightly diferent notation.*" (HINTIKKA, 1962, p. 47-48)

limitações computacionais dos seres humanos e das máquinas “inteligentes”. De fato, o problema da onisciência lógica não interessa de modo algum apenas aos lógicos; tem sido muito discutido também no campo da inteligência artificial, entre outros[15]. A busca pela solução do problema levantado por Hintikka trouxe soluções muito interessantes (algumas delas serão discutidas neste trabalho de tese).

Para compreendermos melhor esse problema, vale a pena retomarmos a discussão sobre onisciência lógica a partir do esquema epistêmico (E-CLOS 1), a saber, $K_a(p \to q) \to (K_a p \to K_a q)$ – que representa o fecho sob implicação material. O que tem sido explicitado até então é que a onisciência lógica é evidenciada quando um dos quatro princípios de fecho, discutidos anteriormente, é satisfeito[16].

É interessante notar que, segundo a noção *folk* de conhecimento, o fecho sob implicação é perfeitamente satisfatível. Para esclarecermos a ideia, observemos o exemplo a seguir, considerando as seguintes proposições:

- K_aP_1: O agente a sabe que a poluição do planeta cresce em níveis elevados.
- K_aP_2: O agente a sabe que o planeta será destruído em um futuro próximo.

Suponha agora que P_1 e a seguinte implicação sejam satisfeitas[17]:

- $K_a(P_1 \to P_2)$: O agente a sabe que a poluição do planeta em níveis elevados implica em uma completa destruição do planeta em um futuro próximo.

Deste modo, temos que ambas K_aP_1 e $K_a(P_1 \to P_2)$ são verdadeiras. A pergunta então é a seguinte:

- É razoável, nesse caso, admitir $K_a(P_1 \to P_2) \to (K_aP_1 \to K_aP_2)$ como um teorema?

Ou seja, admitindo que K_aP_1 e $K_a(P_1 \to P_2)$ sejam verdadeiras, temos também, por implicação lógica (para qualquer indivíduo que goze plenamente de suas

[15] Ver, por exemplo, FAGIN, Ronald; HALPERN, Joseph Y. **Belief, awareness and limited reasoning.** Artificial Intelligence. Vol. 34, p. 39-76, 1988. VARDI, Moshe Y. **On epistemic and logical omniscience.** *In*: TARK'86: Proceedings of the 1986 conference on theoretical aspects of reasoning about knowledge. San Francisco: Morgan Kaufmann Publishers Inc., 1986. p. 293-305.

[16] Onisciência lógica total; fecho sob implicação material; fecho sob implicação válida e o conhecimento de fórmulas válidas.

[17] Deve ser observado, até agora, que nenhuma suposição acerca da verdade de P_2 foi feita. Até então, estamos considerando como satisfeitas apenas P_1 e a implicação de P_1 para P_2.

capacidades racionais), que admitir a verdade de K_aP_2? A noção *folk* de conhecimento sugere que sim. Aquele indivíduo que conhece a verdade de P_1 e, de algum modo, conhece os comportamentos ambientais sabe, sem sombra de dúvida, acerca do perigo que o nosso planeta enfrenta no presente. Esse teorema particular (que pode, sem problema algum, ser estendido a outros contextos) é de fato considerado natural por todos nós, em várias circunstâncias.

O exemplo acima foi somente para mostrar que, neste caso específico e, segundo o senso-comum, o fecho sob implicação parece plausível. Em contrapartida, ele nos leva a pensar então que é o fecho sob implicação válida que parece criar problemas.

Para mostrar a ideia, retomemos o mesmo exemplo, com uma variação:

- K_aP_1: O agente a sabe que a poluição do planeta cresce em níveis elevados.
- K_aP_2: O agente a sabe que o planeta será destruído em um futuro próximo.
- A poluição do planeta em níveis elevados implica em uma completa destruição do planeta em um futuro próximo[18].

A diferença agora é que o agente a não conhece – apesar de que seja o caso – o fato de que a poluição em níveis elevados implicará em uma completa destruição do planeta em um futuro próximo. O agente a apenas sabe que, no momento, a poluição do planeta cresce em níveis elevados.

Temos então a questão: o agente a sabe que o planeta será destruído em um futuro próximo? Não necessariamente. O agente pode falhar em saber que o planeta será destruído em breve, justamente por não conhecer a implicação acima. O exemplo ilustra perfeitamente que, caso o agente saiba que o planeta será destruído em breve, ou ele é logicamente onisciente ou veio a conhecer a implicação a partir de outras fontes. Para conhecer P_2 a partir de P_1, ele precisa conhecer também a implicação $P_1 \rightarrow P_2$. Se ele não a conhece, e não for logicamente onisciente, não há como conhecer P_2 – a não ser que se aumente, por outras fontes, o conhecimento factual de a, no qual a passa então a conhecer também a implicação $P_1 \rightarrow P_2$.

Porém, os argumentos contra a onisciência lógica são também direcionados ao primeiro exemplo, no qual o agente a possui de fato o conhecimento da implica-

[18]Suponha que isso seja uma verdade universal.

ção $P_1 \to P_2$. Argumenta-se que, mesmo que o agente a conheça P_1 e a implicação $P_1 \to P_2$, ele pode não conhecer P_2 por vários motivos, dentre eles:

1. O agente pode simplesmente – por não querer ou não se interessar – não computar todas as consequências lógicas daquilo que ele já sabe.

2. O agente, mesmo gozando normalmente de todas suas capacidades racionais, pode não conseguir computar as consequências lógicas a partir de $K_a(P_1 \to P_2)$ e K_aP_1 (por limitações computacionais ou de tempo[19]).

2.2.5 Senso-comum e onisciência lógica: defesa do esquema epistêmico K em situações específicas

Os argumentos acima nos fazem repensar o modo como o esquema K é entendido. As questões contra as consequências do esquema K parecem ter bastante sentido. Em um primeiro momento, podemos simplesmente abandonar o tipo de lógica no qual o referido esquema seja tido como axioma. De fato, foi o que grande parte dos estudiosos do assunto fez. Hintikka, em seu artigo *Impossible Possible Worlds Vindicated* (1975), propõe uma solução semântica na qual o fecho sob implicação válida não seja satisfeito. Contudo, também é possível mostrar que, em sua abordagem, o fecho sob implicação ainda é derivável. Entraremos em mais detalhes mais à frente. No momento, estamos interessados em mostrar que o esquema K, em algumas situações específicas do senso-comum, pode ser considerado aceitável. Consequentemente, isto sugere que, no âmbito da lógica epistêmica, a problemática da onisciência lógica deva ser tratada seguindo estratégias de aplicação que levam em consideração as especificidades das situações e agentes que se pretende modelar. Logo, as noções gerais de validade e invalidade *simpliciter* de princípios de fecho epistêmico não são de muita utilidade para a lógica epistêmica; esta toma como estratégia de solução de problemas a perspectiva de aplicabilidade da lógica de interesse. Todavia, deixemos esta conclusão de grande peso para o final do capítulo. Porém, novamente, mostrar que um determinado princípio de fecho pode ser aceitável, em circunstâncias específicas, auxiliará esta linha de argumentação. Vejamos, portanto, o caso específico de nosso atual interesse.

Consideremos o mesmo exemplo dado um pouco acima. Hoje em dia, mesmo uma criança sabe das consequências que a poluição trará ao mundo em um futuro

[19]Ainda há outros motivos. Detalhes virão mais à frente.

não muito distante. Ela sabe não necessariamente porque foi dito a ela dessa maneira, mas sim porque já é perfeitamente capaz de seguir implicações lógicas simples. O que queremos então mostrar é que, em vários casos – e em vários contextos diferentes – o agente segue sem problemas as consequências lógicas daquilo que ele conhece, bastando apenas tempo e disposição para computar novamente as novas informações e regras lógicas que passa a conhecer. Assim, o agente pode ainda não conhecer que P_2 é o caso, mas se ele conhece P_1 e $P_1 \rightarrow P_2$, e estiver disposto a analisar todas as consequências do que ele conhece, então o agente poderá, sem maiores problemas, vir a conhecer que P_2 é o caso.

Hintikka, em *Knowledge and Belief* (1962, p. 34), argumentou de maneira similar:

> Por exemplo, o fato de uma sentença da forma (11) [20] ser auto-suficiente não significa que a pessoa referida por a saberá que q tão logo saiba p. Geralmente isso significa meramente que, se ele sabe que p e persegue as conseqüências desse item de conhecimento o bastante, ele também virá a saber que q. Nada é dito se alguém alguma vez o fará[21].

A grande diferença entre a ideia que estamos tentando mostrar e a ideia proposta por Hintikka, é a seguinte: o contexto. Para Hintikka, todo agente capaz de seguir as consequências lógicas do que conhece estaria apto a vir saber que P_2 é o caso. Podemos sugerir, ao invés disso, o seguinte: há contextos em que o mesmo agente aceito por Hintikka é capaz de seguir as consequências lógicas do que conhece, e há outros contextos em que não é capaz disso. Isso vai depender do conhecimento do agente acerca do próprio contexto. Por exemplo, se um agente está ciente de que está numa discussão filosófica, na qual os padrões para aquisição e manutenção de conhecimento são muito altos, ele pode reconhecer ser completamente ignorante acerca da verdade ou falsidade de algumas proposições. Exemplos dessas proposições poderiam ser "não sou uma marionete do gênio maligno cartesiano", "não sou um cérebro numa cuba" etc. Essa seria uma aplicação (contexto) em que agentes pensam filosoficamente sobre hipóteses céticas, e estão cientes disso. Logo, não seria absurdo, aqui, aproximar as noções de contexto e aplicabilidade. Isto é, não seria absurdo afirmar que uma lógica epistêmica, tal

---

[20] $K_a p \rightarrow K_a q$.

[21] *"For instance, tha fact that a sentence of the form (11) is self-sustaining does not mean that the person referred to by a knows that q as soon as he knows that p. Often it means merely that if he knows that p and pursues the consequences of this item of knowledge far enough he will also come to know that q. Nothing is said about whether anybody will ever do so."*

como a de Hintikka, é aplicável a contextos específicos nos quais agentes, com limitações epistêmicas normais, são capazes de efetuar deduções lógicas simples – deduções que envolvam, por sua vez, aplicações bem sucedidas de um determinado princípio de fecho epistêmico. Estes agentes, para os quais esta lógica específica é um modelo, podem satisfazer (ou não) tais e tais princípios de fecho, e assim por diante. Vejamos, na prática, como isto funciona.

Suponha, por exemplo, um agente a qualquer com pós-doutorado em Ecologia. O agente a goza plenamente de suas faculdades racionais, mas não é de modo algum mais capaz racionalmente do que o agente b, que é um pedreiro. Seja X o contexto das coisas relativas ao meio ambiente, e que o presente contexto é o da construção civil em local de dunas móveis. Seja Γ o conjunto de todas as sentenças que o agente a conhece. Entre as consequências de Γ está a seguinte proposição:

- P_3: A construção de prédios em áreas de dunas móveis implicará em sérios impactos ambientais, prejudiciais ao ecossistema local.

Para o agente a, a partir da análise do que já sabe de ecologia, junto com o conhecimento do local em que vive, não é difícil vir a conhecer que P_3 é o caso, e que não deve portanto construir seu imóvel no local indicado[22]. Assim, temos o seguinte:

- O contexto da coisas relativas ao meio ambiente é bem conhecido por a. Em contextos como este, o agente a tem grande facilidade de perseguir as consequências lógicas do que ele conhece sobre o tema. De modo algum a pode ser considerado logicamente onisciente, mas apenas um indivíduo que, nesse contexto, possui um conjunto Γ que facilita a derivação de qualquer consequência lógica relativa à ecologia.

Agora, consideremos o agente b. Seja Γ_1 o conjunto de todas as sentenças que o agente b conhece. O agente b, por sua condição social, não teve a oportunidade, como o agente a, de escolher sua profissão, a qual aprendeu com o próprio pai. Todavia, como foi dito antes, também goza plenamente de todas suas faculdades racionais. Entre as consequências de Γ_1, temos também a proposição P_3. Porém, como não é um estudioso de ecologia como a, os caminhos (e as premissas) para deduzir

[22] P_3 foi escolhida arbitrariamente. A condição que devemos impor é que ela esteja inserida no contexto.

P_3 são bem mais escassos e bem menos evidentes do que aqueles oferecidos pelo conjunto Γ. Sendo assim, apesar de possuir as mesmas capacidades racionais de a, o agente b tem imensa probabilidade de falhar em saber que P_3 é o caso. Logo, o agente b, quando inserido no contexto X, falha ao saber que P_3 é o caso[23]. Todavia se mudarmos o contexto, por exemplo, para a construção e manutenção de um imóvel, é extremamente provável que o agente b conheça mais consequências daquilo que já lhe é conhecido. Assim, mesmo ao considerarmos que tanto Γ quanto Γ_1 derivam:

- P_4 O alicerce não deve ser paralelo ao solo.

É bem mais provável que b consiga mais rápido do que o agente a, através da análise lógica do que já conhece, vir a conhecer que P_4 seja o caso.

Tudo isso foi útil para mostrar que, mesmo dois agentes que possuem capacidades racionais iguais não são capazes de perseguir, com os mesmos resultados, as consequências lógicas do que conhecem explicitamente. Dizemos, então, que o agente é capaz de vir a conhecer as consequências lógicas que se seguem a partir de seu conhecimento, com maior ou menor eficiência, dependendo do contexto no qual ele esteja inserido, e as proposições sobre as quais irá raciocinar. Daí a importância, para qualquer lógica epistêmica, definir que tipo de agente (ou agentes) ela pretende modelar; além disso, que tipo de noção de conhecimento ou crença ela pretende capturar. Reduzir princípios de fecho a categorias permanentes de "princípios epistêmicos válidos" ou "princípios epistêmicos" inválidos é desconsiderar a especificidade de cada lógica – assim entendemos.

Esta posição tem um propósito útil. Não podemos aceitar o esquema epistêmico K de maneira absoluta, isto é, aplicado a qualquer agente e a qualquer contexto. Porém, uma lógica que tem como axioma este esquema pode muito bem ser considerada uma sub-lógica da lógica de outrem, na qual é perfeitamente plausível (devido ao bom conhecimento do contexto por parte dos agentes) se perseguir as consequências lógicas daquilo que se conhece. A lógica que teria como axioma o esquema K seria, portanto, uma lógica epistêmica de agentes e contextos particulares, na qual K seria uma fórmula válida exatamente nos contextos afins aos agentes, contextos estes que lhes permitem, por sua vez, seguirem as consequências lógicas daquilo que conhecem.

[23]Suponha que os dois, tanto a quanto b são pessoas que costumam sempre revisar aquilo que conhecem, e seguir as consequências lógicas de seu conhecimento factual.

Continuando a discussão sobre a onisciência lógica, retornamos ao ponto inicial, que é a solução que Hintikka propõe em *Knowledge and Belief*. Mais tarde, a discussão é retomada por ele em seu artigo *Impossible Possible Worlds Vindicated* (1975), que também será alvo de análise.

2.2.6 Discussão sobre a onisciência lógica em *Knowledge and belief*

Retomando o que foi dito no início deste capítulo, Hintikka identifica, logo cedo (1962, p. 30), o problema da onisciência lógica em sua abordagem:

> Por meio das minhas regras, é visível que (11) [$K_a p \rightarrow K_a q$] é válida tão logo p implique q em nossa lógica proposicional ordinária. Mas é claramente inadmissível inferir 'Ele sabe que q' de 'Ele sabe que p' baseando-se somente no fato de que q segue logicamente de p, pois a pessoa em questão pode falhar em saber que p implica q, particularmente se p e q são sentenças relativamente complicadas[24].

Podemos, então, fazer a seguinte questão: e se por acaso o agente de fato conhecesse a implicação $p \rightarrow q$? O que diríamos? O agente conheceria q? Observem que Hintikka aponta para o fato de que o agente não conhece q por não conhecer a implicação $p \rightarrow q$. Assim, para Hintikka, esse agente pode ser considerado logicamente onisciente se $K_a q$ é obtida a partir de $p \rightarrow q$ e $K_a p$. Isso nos dá a impressão de que, para Hintikka, a onisciência lógica surge a partir do fecho sob implicação válida. O que dizer então do fecho sob implicação? Observando a citação acima, parece-nos que Hintikka considera perfeitamente natural a obtenção de $K_a q$ a partir de $K_a(p \rightarrow q)$ e $K_a p$. Ao contrário do fecho sob implicação válida, o fecho sob implicação nem chega a ser questionado por Hintikka em *Knowledge and Belief*. Ou seja, o esquema K continua um teorema de seu sistema. Isso acontece devido a algo muito simples: existem vários – e não apenas um – motivos para a falha em onisciência lógica por parte dos agentes "reais". Sendo assim, mesmo conhecendo a implicação $p \rightarrow q$, o agente pode falhar em saber que q é o caso. A tese discutida por Hintikka lida com o fato de o agente não conhecer a implicação $p \rightarrow q$. Assim, ela capta uma forma específica de falha em onisciência lógica: aquela em que o agente não conhece aquilo que chamamos de "implicação relevante" (Ver capítulo 1).

[24]"*By means of my rules, it is readily seen that (11) [$K_a p \rightarrow K_a q$] is valid as soon as p logically implies q in our ordinary propositional logic. But it is clearly inadmissible to infer 'He knows that q' from 'He knows that p' solely on the basis of the fact that q follows logically from p, for the person in question may fail to see that p entais q, particulary if p and q are relatively complicated statements.*"

Como já podemos perceber a esta altura, diferentes lógicas epistêmicas captam diferentes falhas em onisciência lógica, validam ou invalidam diferentes princípios de fecho e modelam o conhecimento de diferentes tipos de agentes. Já discutimos, no primeiro capítulo, algumas das razões pelas quais agentes podem falhar em satisfazer a propriedade de onisciência lógica. Entre essas razões, encontramos: (1) recursos limitados, (2) ausência de consciência dos conceitos relevantes e (3) preconceitos. Além destas, como mostram Fagin & Halpern (1998, p. 40-41), podemos destacar:

(4) Atenção desconexa[25]. Às vezes, as pessoas não conseguem prestar atenção a todos os temas simultaneamente. Quando dizemos, por exemplo, que sabemos que uma proposição p qualquer é o caso, estamos dizendo que, em todas as alternativas epistêmicas que consideramos, acontece de p ser verdadeira[26]. A partir de cada alternativa epistêmica considerada – na qual p é o caso – podemos extrair certas conclusões lógicas. Por sermos racionalmente limitados, e não prestarmos uma atenção detalhada a cada uma dessas alternativas, podemos extrair conclusões de umas que são inconsistentes com certas conclusões de outras, já que cada alternativa epistêmica pode ser muito diferente uma da outra. Assim, apesar do fato de cada alternativa epistêmica considerada ser individualmente consistente – isto é, não ocorrer nela que tenhamos $p \wedge \neg p$ – as conclusões que extraímos de cada uma delas podem nos levar a sustentar proposições inconsistentes – do tipo $p \wedge \neg p$, por exemplo.

Estas são apenas algumas razões pelas quais lógicos ou epistemólogos podem argumentar contra a propriedade de onisciência lógica. Quais dessas propriedades captar, ou que tipo de agentes modelar é uma opção que depende da aplicação da lógica epistêmica de interesse. Hintikka, por exemplo, com o intuito de modelar agentes com capacidades epistêmicas normais, observou que deveria, de algum modo, mostrar que os agentes de sua lógica não eram logicamente oniscientes. Ele tentou fazer isto através da seguinte estratégia: a alteração da noção clássica de "mundo possível". Como veremos, abandonar a noção padrão de mundo possível não é a única opção disponível para quem quer invalidar certos princípios de fecho em um sistema de lógica epistêmica (FAGIN *et al*, 2003, p.336-337). No entanto, é certo que tal estratégia não deixa de ser útil em alguns casos. Vejamos, a seguir, o

[25]A expressão "atenção desconexa" não está presente no trabalho dos autores supracitados. O termo foi criado com o intuito de tornar esta exposição mais didática.

[26]Como vimos, esta é a definição padrão de conhecimento em lógica epistêmica.

caminho tomado por Hintikka e os seus famosos "impossíveis mundos possíveis".

2.2.7 A solução proposta em *Impossible Possible Worlds Vindicated*

A solução do artigo *Impossible Possible Worlds Vindicated* (HINTIKKA, 1975) mostra-se interessante o suficiente para gerar bastante discussão. Apesar de, a princípio, os *impossíveis mundos possíveis* parecerem entidades um tanto esquisitas, a ideia se mostra muito plausível quando comparada à semântica de modelos não-padrão para a lógica clássica de primeira ordem, de Rantala (1975)[27]. Em seu artigo de 1975, Hintikka tenta novamente mostrar que sua abordagem de 1962 está livre de qualquer comprometimento com a onisciência lógica.

Antes de expormos seu argumento, achamos conveniente mostrar, da mesma maneira que ele o fez (HINTIKKA, 1975, p. 475), o que supostamente levaria sua abordagem ao comprometimento com a onisciência lógica. Para isso, escolhemos, primeiro, estabelecer o modelo de Hintikka em uma notação familiar, caracterizando um modelo semântico para as noções de conhecimento com base nas estruturas de Kripke (1959).

Semântica padrão de mundos possíveis para a noção de conhecimento

A intuição por trás da semântica padrão de mundos possíveis é a de que, além do nosso estado real de coisas (o mundo "real"), existe um outro número de estados de coisas. Esses outros estados de coisas também são conhecidos como "mundos possíveis". Formalizemos então essa ideia. Lembremos, contudo, que tal ideia sofre aqui uma pequena modificação, pois estamos tratando de lógica epistêmica. Sendo assim, ao invés de utilizarmos o termo "mundo possível", convém mais a terminologia "alternativa epistêmica". Deste modo, dizemos que, além de considerar o atual estado de coisas, um agente pode considerar vários outros estados alternativos ao atual. Daí a utilização do termo.

Iniciemos então a construção da semântica, começando pela linguagem.

Definição 2.1. *(Linguagem) Seja $\mathscr{L}$ o símbolo para a nossa linguagem. A linguagem $\mathscr{L}$ é formada a partir de:*

[27] RANTALA, Veikko. Urn models: A new kind of non-standard model for first-order logic. **Journal of Philosophical Logic.** Vol. 4, n. 3, p. 455-474, ago. 1975.

1. *Símbolos lógicos:* $\neg$, $\vee$, $\wedge$, $\rightarrow$;

2. *Símbolo epistêmico:* K;

3. *Um conjunto infinito enumerável* $\{p_1, p_2, ..., p_k\}$ *de letras proposicionais;*

4. *Um conjunto infinito enumerável* $\{a, b, c, ..., a_1, b_1, c_1, ..., a_k, b_k, c_k\}$ *de variáveis representando agentes epistêmicos;*

5. *Símbolos auxiliares (,) (Parênteses).*

De posse dos símbolos da linguagem, podemos então definir o que é uma fórmula:

Definição 2.2. *(fórmulas proposicionais epistêmicas): O conjunto das fórmulas proposicionais epistêmicas é especificado pelas seguintes regras:*

1. *Toda letra proposicional é uma fórmula;*

2. *Se* α *é uma fórmula, então* $\neg\alpha$ *também é uma fórmula;*

3. *Se* α *e* β *são fórmulas, e* $\circ$ *é um operador binário,* $\alpha \circ \beta$ *também é uma fórmula;*

4. *Se* α *é uma fórmula, e* a *um símbolo de agente epistêmico, então* $K_a\alpha$ *é uma fórmula.*

Definição 2.3. *(símbolos definidos) São símbolos definidos os seguintes:*

1. $\alpha \leftrightarrow \beta =_{def} (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$.

2. $P_\sigma\alpha =_{def} \neg K_\sigma \neg\alpha$.

A seguir, mais algumas definições importantes que nos permitirão visualizar o comportamento semântico do esquema K e a solução de Hintikka em um âmbito formal.

Definição 2.4. *(Estrutura semântica): Uma estrutura* $\mathscr{E}$ *consiste em um conjunto não-vazio* W*, no qual seus membros* $w_1, ..., w_K$ *são chamados de alternativas epistêmicas, e uma relação binária* R *em* W*, chamada de relação de acessibilidade. Deste modo, uma estrutura é um par* $\mathscr{E} = \langle W, R \rangle$.

Para que possamos definir "validade", necessitamos antes de duas definições, a saber, a definição de "modelo" e de "verdade em um modelo". Um *modelo* permite mostrar quais fórmulas são verdadeiras em quais alternativas epistêmicas.

Definição 2.5. *(modelo)*[28]*: Um modelo $\mathscr{M}$ é uma tripla $\langle W,R,\Vdash\rangle$, na qual $\langle W,R\rangle$ é uma estrutura e $\Vdash$ uma relação entre alternativas epistêmicas e letras proposicionais. Se $w \Vdash p$ ocorre, dizemos que a fórmula p é o caso em w. Se $w \Vdash p$ não ocorre, simbolizamos $w \nVdash p$ e dizemos que a fórmula p não é o caso em w.*

Definição 2.6. *(verdade em um modelo): Seja $\mathscr{M} = \langle W,R,\Vdash\rangle$ um modelo. A relação $\Vdash$ é estendida a fórmulas arbitrárias da seguinte maneira. Para cada $w \in W$:*

1. *$w \Vdash \neg\alpha$ se, e somente se, $w \nVdash \alpha$;*

2. *$w \Vdash \alpha \wedge \beta$ se, e somente se, $w \Vdash \alpha$ e $\Vdash \beta$;*

3. *$w \Vdash \alpha \vee \beta$ se, e somente se, $w \Vdash \alpha$ ou $w \Vdash \beta$;*

4. *$w \Vdash \alpha \to \beta$ se, e somente se, $w \Vdash \alpha$ então $w \Vdash \beta$;*

5. *$w \Vdash K_a\alpha$ se, e somente se, para todo $w' \in W$, se, wRw' então $w' \Vdash \alpha$;*

Agora, com base em tudo que construímos até então, partimos para a definição mais importante, que é a de fórmula L-válida. Com ela, podemos verificar a validade (ou invalidade) de todas as regras estabelecidas por Hintikka[29], incluindo o tão discutido esquema K. Nesse estilo de semântica, a solução que Hintikka propõe para a onisciência lógica é fácil de ser compreendida; além disso, é possível também encontrarmos alguns pontos controversos acerca desta solução. Continuemos então apresentando a definição.

Definição 2.7. *Se L é uma coleção de estruturas, uma fórmula α é L-válida se ela é válida em todas as estruturas de L.*

[28]Se quisermos ser mais rigorosos, temos de utilizar o termo "modelo proposicional epistêmico".

[29]Na verdade, o que fazemos ao utilizar essa semântica é testar a validade de esquemas que, se prestarmos atenção, representam no nosso modelo as mesmas regras de Hintikka. Um exemplo disso é o esquema epistêmico T: $K_a p \to p$, que, ao observarmos, é a regra $(C.K)$ traduzida para esse estilo de semântica.

Explicitação do argumento

Retomando a ideia de Hintikka (1975, p. 475), observamos que:

1. Uma sentença da forma $K_a p$ é o caso em um mundo $w \in W$ se e somente se p é o caso em todo $w' \in W$, wRw'. Informalmente, uma sentença da forma 'a conhece que p' é o caso se, e somente se, p é o caso em todas as a-alternativas epistêmicas para w. Ou seja, em todas as circunstâncias possíveis que são compatíveis com tudo o que a conhece em w. Aqui, as alternativas epistêmicas podem ser pensadas de várias maneiras como, por exemplo, "estado de coisas", "situações", "curso de eventos" etc.
2. Nessa abordagem, uma sentença é logicamente verdadeira quando ela é o caso em todas as alternativas epistêmicas de um modelo.
3. A falha da onisciência lógica deve se dar da seguinte maneira: seja $\langle W, R, \Vdash \rangle$ um modelo qualquer. Suponhamos um agente qualquer a. Sejam p e q duas sentenças quaisquer. Suponha então que $w \Vdash K_a p$ (para um w qualquer de μ) e que $p \rightarrow q$ é logicamente verdadeira. Porém, $w \nVdash K_a q$.

Hintikka aponta que a crítica acerca de sua formulação da lógica epistêmica repousa sobre o fato de que 1–3 são propriedades incompatíveis. O problema então nos dá duas alternativas imediatas: reconhecer a incompatibilidade e tentar resolvê-la, ou então mostrar que de fato não há incompatibilidade alguma. A opção preferida por ele é a segunda.

A ideia então é mostrar que a incompatibilidade entre 1-3 só ocorre quando a seguinte propriedade adicional é satisfeita:

4. Toda alternativa epistêmica é logicamente possível (*simpliciter*). Ou seja, todo $w \in W$ é logicamente possível.

Aceitando 4 juntamente com 1, 2 e 3, fica fácil perceber a contradição. Suponha um modelo qualquer $\mathscr{M} = \langle W, R, \Vdash \rangle$ baseado em uma estrutura qualquer $\mathscr{E} = \langle W, R \rangle$. Suponha, então, que o modelo em questão satisfaz a propriedade 2. Sendo assim, considere duas sentenças p e q e um agente qualquer a, de modo que $w \Vdash K_a p$ (para um w qualquer de $\mathscr{M}$) e $p \rightarrow q$ é logicamente verdadeira (isto é, $\vDash p \rightarrow q$), mas $w \nVdash K_a q$. Ora, se $w \nVdash K_a q$, então existe uma alternativa epistêmica w'

em W, wRw', na qual q não é o caso; ou seja, $w' \nVdash q$. Já foi assumido por nós que $K_a p$. Sendo assim, p é o caso em em todo w', wRw'. Isto é, $w' \Vdash p$, para todo $w' \in W$, wRw', do modelo $\mathscr{M}$. Se aceitamos a propriedade 4, então toda alternativa epistêmica é logicamente possível. Sendo assim, w' é logicamente possível. Como w' é logicamente possível e $p \rightarrow q$ é logicamente verdadeira, $w' \Vdash p \rightarrow q$. Ora, se $w' \Vdash p$ e $w' \Vdash p \rightarrow q$, pela definição 2.6-4 temos que $w \vDash q$. Porém, como já havíamos aceitado antes, $w' \nVdash q$. Eis então nossa contradição. A adição da hipótese 4 é o que torna a contradição possível.

Porém, quais são as razões que nos levam a rejeitar 4? Será que podemos conceber alguma alternativa epistêmica que não seja uma alternativa logicamente possível? Essa ideia é coerente? Ou é apenas um subterfúgio formal de Hintikka para mostrar que os agentes de sua lógica não são logicamente oniscientes? O que podemos verificar nessa solução é o seguinte:

1. O argumento apresentado nos permite conceber $\nVdash (K_a p \wedge p \rightarrow q) \rightarrow K_a q$. Porém, quando $K_a p$ e $K_a(p \rightarrow q)$ são satisfeitas, $K_a q$ também o é. Assim, como já foi comentado anteriormente, o esquema epistêmico K continua válido.

2. A negação de 4 nos leva a aceitar o que o próprio Hintikka irá chamar de "impossíveis mundos possíveis". Sendo assim, a princípio, a solução de Hintikka em favor dessas entidades parece complicar mais ainda o problema, ao invés de resolvê-lo.

Veremos, mais à frente, a ideia que permeia essa argumentação contra a hipótese 4. Antes disso, porém, achamos interessante mostrar o comportamento formal de uma semântica que não aceita a propriedade 4. Sendo assim, poderemos verificar formalmente aquilo que viemos mostrando, como por exemplo, a invalidade de $(K_a p \wedge p \rightarrow q) \rightarrow K_a q$ e a validade do esquema epistêmico K.

Validade e invalidade de fórmulas importantes

Sabemos que a solução proposta por Hintikka nos permite demonstrar $\nVdash (K_a p \wedge p \rightarrow q) \rightarrow K_a q$ a partir da negação da hipótese 4. Deste modo, os impossíveis mundos possíveis inviabilizam (E-CLOS 6), isto é, o fecho sob implicação válida. Essa demonstração segue da seguinte maneira:

Suponha um modelo qualquer $\mathscr{M} = \langle W, R, \Vdash \rangle$ baseado em uma estrutura qualquer $\mathscr{E} = \langle W, R \rangle$. Suponha uma alternativa qualquer $w \in W$ tal que $w \Vdash K_a p$. Considere também que $p \rightarrow q$ é uma verdade lógica. Ora, se $w \Vdash K_a p$, então $w' \Vdash p$, para todo $w' \in W$, wRw'. A negação da propriedade 4 nos permite inserir na prova que, entre todas as alternativas epistêmicas a w, existe ao menos uma que é um "impossível mundo possível". Sendo assim, entre todos os $w' \in W$, existe ao menos um w' no qual certas verdades lógicas não valem. Deste modo, podemos supor um w' tal que $w' \nVdash p \rightarrow q$. Neste mesmo w', sabemos que p é o caso, pois o foi considerado pelo próprio agente a. Temos então que $w' \Vdash p$ e $w' \nVdash p \rightarrow q$. Assim, pela definição 2.6-4, $w' \nVdash q$. Em seguida, aplicando a definição 2.6-5 em $w' \nVdash q$, temos que $w \nVdash K_a q$. Logo, $\nVdash (K_a p \wedge p \rightarrow q) \rightarrow K_a q$.

Devemos observar uma propriedade interessante. Neste mesmo modelo, temos $w \Vdash q$, pois $w \Vdash p$ e $\vDash p \rightarrow q$. Sendo assim, apesar de q ser o caso em w, o agente a não o sabe. Mesmo que q seja uma consequência lógica do que ele conhece, e mesmo que q seja o caso, o agente a falha em saber de q. Em contrapartida:

Suponha um modelo qualquer $\mathscr{M} = \langle W, R, \Vdash \rangle$ baseado em uma estrutura qualquer $\mathscr{E} = \langle W, R \rangle$. Suponha uma alternativa qualquer $w \in W$ tal que $w \Vdash K_a p$ e $w \Vdash K_a(p \rightarrow q)$. Assim, pela definição 2.6-5, $w' \Vdash p$ e $w' \Vdash p \rightarrow q$, para todo $w' \in W$, wRw'. Deste modo, se $w' \Vdash p$ e $w' \Vdash p \rightarrow q$, por 2.6-4 temos que $w' \Vdash q$, para todo $w' \in W$. Assim, por 2.5-5, $w \Vdash K_a q$. Logo, $w \Vdash K_a(p \rightarrow q) \rightarrow (K_a p \rightarrow K_a q)$. Como w é uma alternativa arbitrária, temos que $w \Vdash K_a(p \rightarrow q) \rightarrow (K_a p \rightarrow K_a q)$ é o caso em todo w do modelo em questão. O modelo e a etrutura também são arbitrários. Daí, pela definição 2.7, a fórmula $K_a(p \rightarrow q) \rightarrow (K_a p \rightarrow K_a q)$ é válida.

Esta foi, portanto, a demonstração da validade do esquema epistêmico K a partir da presente semântica. Deste modo, o fecho sob implicação ainda é uma propriedade válida na lógica de Hintikka, mesmo em uma semântica com impossíveis mundos possíveis. Mesmo considerando que w' fosse um impossível mundo possível, isso não afetaria a validação do esquema K. Ao supormos $w \vDash K_a p$ e $w \vDash K_a(p \rightarrow q)$, estamos aceitando que p e $p \rightarrow q$ são o caso, independentemente do w' em questão. Devemos atentar para o significado do operador K. O agente a conhece algo quando esse algo é de fato o caso em todas as alternativas epistêmicas consideradas por ele. Aqui, não importa se w' é ou não logicamente possível. O que importa é que, independentemente de qual w', p e $p \rightarrow q$ são sempre o caso em w', dado que o agente conhece p e $p \rightarrow q$.

Em seguida, temos mais uma fórmula que traz problemas à solução de Hintikka, a saber, o esquema epistêmico T, $K_a p \rightarrow p$. Na semântica original de Hintikka, o esquema $K_a p \rightarrow p$ é representado através da regra $(C.K)$, proposta e defendida por Hintikka em *Knowledge and Belief*. Sendo assim, em acordo com a semântica de Hintikka, nossa semântica também deve validar o esquema T. A validação do esquema T segue sem problema se aceitarmos a propriedade 4. Porém, Hintikka a nega com o propósito de evitar a onisciência lógica. O que mostraremos, então, é que uma semântica com impossíveis mundos possíveis prejudica a aceitação de uma regra defendida por ele mesmo – a regra $(C.K)$.

Primeiramente, aceitemos a validade de $K_a p \rightarrow p$. Suponha um modelo qualquer $\mathscr{M} = \langle W, R, \Vdash \rangle$ baseado em uma estrutura qualquer $\mathscr{E} = \langle W, R \rangle$. Suponha uma alternativa qualquer w tal que $w \Vdash K_a p$. Ora, se $K_a p \rightarrow p$ é válida e $w \Vdash K_a p$, então $w \Vdash p$. Esse não é o propósito da prova. Lembremos que a solução de Hintikka nega a propriedade 4. Sendo assim, considere ainda a hipótese $w \Vdash K_a p$. Pela definição 2.6-5, temos que $w' \Vdash p$, para todo $w' \in W$, wRw'. Novamente, a negação da hipótese 4 nos permite admitir que, dentre todos os w', existe ao menos um no qual certas verdades lógicas não valem (ou todas, dependendo do w'). Considere agora a seguinte verdade lógica: $p \rightarrow p$. Podemos, pela negação de 4, aceitar então um w', wRw', tal que $w' \nVdash p \rightarrow p$. Ora, se $w' \nVdash p \rightarrow p$, então, pela definição 2.6-4, $w' \Vdash p$ e $w' \nVdash p$. Devemos observar que o argumento de Hintikka se aplica de uma maneira geral. Não há qualquer restrição sobre qual verdade lógica devemos ou não contestar. Sendo assim, não há problema em admitir um w' no qual a verdade lógica $p \rightarrow p$ não é o caso. Considerando, então, que $w' \nVdash p$ e wRw', obtemos $w \nVdash K_a p$. Assim, pela definição 2.6-1, $w \Vdash \neg K_a p$. Ora, se $w \Vdash K_a p$ e $w \Vdash \neg K_a p$, por 2.6-2 temos que $w \Vdash K_a p \wedge \neg K_a p$. Eis então uma contradição a partir da simples assunção de $K_a p$.

Mas o resultado final não para por aqui. Foi aceito, de antemão, a validade de $K_a p \rightarrow p$. Considerando que $w \Vdash K_a p \wedge \neg K_a p$ e que estamos em uma lógica clássica, temos como consequência $w \Vdash \neg p$[30]. Logo, aceitando a validade de $K_a p \rightarrow p$ e a hipótese $w \Vdash K_a p$, chegamos não somente a $w \Vdash p$, mas também a $w \Vdash \neg K_a p$ e $w \Vdash \neg p$ – isso tudo se seguirmos o argumento de Hintikka e negarmos a propriedade 4.

Não sabemos qual o resultado pior, se é a negação da regra $(C.K)$ ou a obtenção de $\neg K_a p$ a partir de $K_a p$[31].

[30] A partir da fórmula proposicionalmente válida $p \rightarrow (\neg p \rightarrow q)$.

[31] Este resultado, apresentado na XVII Semana de Filosofia, CCHLA – UFRN, 2007, mostra que a

A ideia por trás dos "impossíveis mundos possíveis"

O que acabamos de mostrar é uma mera consequência formal da aceitação do argumento proposto por Hintikka. Apesar de nos levar a resultados formais bastante inconvenientes, a solução dos "impossíveis mundos possíveis", quando examinada com atenção, não parece tão estranha. É claro que, na discussão geral sobre o tema, encontramos pontos positivos e negativos. Com relação ao problema da onisciência lógica, Lipman escreve:

> Felizmente, existe uma solução simples – até óbvia – para o problema. Se algum dos mundos que o agente concebe como sendo possível não é logicamente consistente, então a cadeia de raciocínio acima é quebrada. Se o agente concebe um mundo no qual $\varphi \to \psi$ é verdadeiro, φ é verdadeiro, mas ψ é falso, então aprender φ não leva o agente a reconhecer que ψ é verdadeiro, mesmo se o agente já sabe que $\varphi \to \psi$ é verdadeiro[32]. (LIPMAN, 1994, p. 183)

Em contrapartida:

> A dificuldade dessa solução é, infelizmente, um tanto óbvia: o que deveríamos assumir com relação aos impossíveis mundos possíveis? Colocando de modo diferente, exatamente quais lógicas não-padrão deveríamos usar para descrever o raciocínio de agentes reais? Já está bem claro o que o "raciocínio perfeito" implica; mas não está claro de modo algum como dar uma formulação do "raciocínio imperfeito"[33]. (LIPMAN, 1994, p. 183)

Apesar de tudo – como dissemos anteriormente – quando analisados com atenção, os "impossíveis mundos possíveis" não parecem tão estranhos. Mas então, o que seria de fato um "impossível mundo possível"? Seja w uma alternativa epistêmica que descreve o estado de coisas atual. Uma alternativa epistêmica $w^{\prime}$ a w consiste em uma contingência que é deixada aberta a partir de qualquer coisa que a saiba em w. Os "impossíveis mundos possíveis" são, nada mais nada menos, do que contingências aparentes consideradas pelo agente. Este agente, por ser logicamente limitado, pode ser vítima de **atenção desconexa**; nessas circunstâncias,

---

aceitação dos "impossíveis mundos possíveis" leva à trivialização do sistema – caso não sejam feitas certas ressalvas às definições.

[32] *"Fortunalety, there is a simple – even obvious – solution to the problem. If some of the worlds the agent conceives of as possible are not logically consistent, then the chain of reasoning above is broken. If the agent conceives of a world in which $\varphi \to \psi$ is true, φ is true, but ψ is false, then learning φ does not lead the agent to recognize that ψ is true, even if he already knows that $\varphi \to \psi$ is true."*

[33] *"The difficulty with this solution, unfortunately, is also quite obvious: what should we assume about the impossible possible worlds. Put diferently, exactly which nonstandard logic should we use to describe the reasoning of real agents? It is quite clear what "perfect reasoning" entails; it is not at all obvious how to give a precise formulation of "imperfect reasoning"."*

o referido agente pode não conseguir prestar atenção a todos os temas ao mesmo tempo. Daí, uma determinada contingência aparente pode conter informações conflitantes com outra contingência qualquer, sem que o agente as perceba. Hintikka argumenta que, sustentar a propriedade 4 requer que aceitemos a plena capacidade do agente de eliminar todas as contingências aparentes, o que não é o caso (HINTIKKA, 1975, p. 64). A aceitação de 4, portanto, implica que o agente só concebe alternativas objetivamente possíveis, excluindo toda e qualquer contingência que contenha contradições lógicas. Logo, a lógica epistêmica de Hintikka poderia, sem problemas, ser um modelo lógico que capta a falha de onisciência lógica por razões de atenção desconexa.

A solução de Hintikka é bem sucedida na medida em que apenas a condição da falha de onisciência lógica número 4 (atenção desconexa) é satisfeita. Isto é, o agente não consegue prestar atenção a todos temas ao mesmo tempo. Por esta razão, ele considera certas contingências aparentes, que são incompatíveis com outras que ele também considera. Estas contingências rejeitam certas verdades lógicas, e por isso impedem-no, às vezes, de seguir as consequências lógicas daquilo que já é conhecido pelo referido agente. Para as demais condições de falha de onisciência lógica, a solução de Hintikka dos "impossíveis mundos possíveis" não é apropriada.

A grande dificuldade de se propor uma solução para o problema da onisciência lógica é a seguinte: satisfazer as várias condições de falha de onisciência lógica ao mesmo tempo. Sendo assim, quanto mais condições forem satisfeitas, melhor será a lógica epistêmica. Será mesmo?

Este argumento pode, de certo modo, ser enfraquecido. Nem sempre os agentes falham em ser logicamente oniscientes por conta de todos esses motivos que apresentamos. De fato, os motivos podem ser vários, mas não necessariamente todos ao mesmo tempo. Um agente pode dominar todos os recursos computacionais ao seu dispor, mas falhar em saber p pelo fato de p ser um conceito que lhe é completamente alheio. É claro que, quanto mais condições forem satisfeitas, melhor para trabalharmos em lógica epistêmica; isso por conta da diversidade de agentes e situações diferentes que poderemos modelar. Contudo, também podemos selecionar lógicas, agentes e situações particulares para construir modelos lógico-epistêmicos específicos, para modelar situações e agentes particulares, segundo nossa conveniência. Como pudemos observar, seria falso afirmar que a lógica epistêmica de

Hintikka, acrescida dos impossíveis mundos possíveis, é incapaz de modelar qualquer circunstância de falha em onisciência lógica.

Novamente: o sucesso ou insucesso de uma lógica epistêmica está diretamente relacionado a sua aplicabilidade. Vale ressaltar, no entanto, que o critério da aplicabilidade não precisa ser fixo. Porém, no que concerne ao problema da onisciência lógica e aos princípios de fecho epistêmico, a literatura tem mostrado o seguinte: o critério de aplicabilidade é, certamente, essencial para a avaliação de uma lógica epistêmica que tenta lidar com invalidação de princípios de fecho, bem como a propriedade de onisciência lógica.

Minha sugestão, portanto, é que esta mesma estratégia seja absorvida pelos epistemólogos que lidam com o problema da (in)validade de princípios de fecho epistêmico. Para isto, seguirei mostrando mais algumas soluções formais para o problema da onisciência lógica. Continuaremos vendo, durante este capítulo, sempre a mesma tendência: cada lógica epistêmica capta a invalidade de certos princípios de fecho epistêmico segundo seus interesses. Daí, nem sempre estas lógicas irão invalidar os mesmos princípios de fecho; nem sempre irão captar as mesmas razões de falha em onisciência lógica.

Para continuarmos a discussão, convém apresentar abordagens que satisfaçam outras condições de falha de onisciência lógica. A próxima solução, na seção seguinte, será a *lógica das crenças explícitas e implícitas*, de Levesque[34].

2.3 Crenças implícitas e explícitas, e a introdução da noção de consciência

Nesta seção, apresentaremos três abordagens que lidam com o problema da onisciência lógica fazendo uma distinção entre crenças implícitas e explícitas. Além disso, há também a introdução da noção de "consciência". As propriedades mais importantes de cada abordagem serão demonstradas e comentadas. Finalmente, ao final de cada abordagem, discutimos os resultados.

A intenção do procedimento aplicado em toda esta seção é a de também deixar transparente a ideia de complementaridade que cada abordagem comporta com relação a outra. Não faremos, pois, exclusão de uma abordagem em prol de

[34]LEVESQUE, Hector J. **A logic of implicit and explicit belief**. *In*: Proceedings of the national conference on artificial intelligence. Austin: AAAI Press, 1984, p. 198-202.

outra. Deste modo, seremos coerentes com a tese da avaliação de princípios de fecho segundo a aplicabilidade da lógica de interesse. Insistiremos, sempre, que tal abordagem seja também aquela adotada na epistemologia informal. Assim, sugerimos a restrição dos conceitos de validade e invalidade na discussão filosófica acerca do fecho epistêmico. Esta restrição dos conceitos de validade e invalidade é, como se pôde perceber até então, bastante comum em lógica epistêmica, dado que cada sistema epistêmico (in)valida conjuntos de fórmulas segundo os propósitos de seu próprio modelo.

De fato, uma das ideias que sustentamos é a de que cada abordagem da lógica epistêmica tem seus pontos fortes e fracos, mas nenhuma é superior a outra. Mesmo que haja uma lógica epistêmica (ou da crença) que satisfaça todas as condições de falha onisciência lógica apresentadas, ainda assim não se pode concluir, necessariamente, que essa lógica é superior às demais, que são menos abrangentes. A questão que decorre daí é a mesma que aparece na filosofia da ciência: quais são os critérios para dizer quando uma teoria é melhor do que outra? Essa questão está além do âmbito deste trabalho, e portanto não será examinada. Fica, contudo, um interessante problema a ser investigado em outro momento: existem critérios suficientes capazes de decidir quando uma lógica epistêmica é mais adequada que outra para solucionar o problema da onisciência lógica, em um âmbito geral?

Retornando ao tema central desta seção, concentremo-nos em apresentar e analisar as abordagens que se seguem.

2.3.1 Lógica das crenças implícitas e explícitas de Levesque

A lógica de Levesque propõe um modelo lógico para a noção de crença, ao invés de conhecimento. Contudo, não há problemas, já que a primeira pode, no que se refere ao método formal, ser tratada similarmente à segunda. O propósito fundamental dessa lógica é basicamente o mesmo: invalidar as propriedades (princípios de fecho) que levam à onisciência lógica. Em Levesque, encontramos certas diferenças em relação ao modelo clássico. Em lugar de apenas um operador para ‘crença’, temos dois operadores:

1. Operador B: crenças explícitas.

2. Operador L: crenças implícitas.

Um detalhe importante a ser observado na lógica de Levesque é que as crenças implícitas incluem todas as consequências lógicas das crenças explícitas.

A linguagem é construída de modo usual. A diferença encontrada na lógica de Levesque é que não há reiteração de modalidades. Sua linguagem é construída de modo que nem B nem L aparecem sob o escopo um do outro. Como símbolos proposicionais primitivos, temos $\wedge$ e $\neg$. Os símbolos proposicionais $\rightarrow$, $\vee$ e $\leftrightarrow$ podem ser definidos também de modo usual. As cláusulas para a construção de fórmulas também não constituem grande diferença da lógica epistêmica clássica:

Definição 2.8. *Se φ é uma fórmula proposicional, então $B\varphi$ e $L\varphi$ também são fórmulas*[35].

Levesque não sente a necessidade de assumir a constante $\top$ (*true*). A seguir, temos a definição de estrutura.

Definição 2.9. *Uma estrutura semântica η em Levesque consiste de quatro elementos: $\eta = (S, \mathscr{B}, T, F)$. Cada elemento de η:*

1. *S: conjunto primitivo de "situações";*
2. *$\mathscr{B}$: subconjunto de S. $\mathscr{B}$ representa as situações alternativas à situação atual;*
3. *T e F: funções que vão de ϕ (conjunto das proposições primitivas) aos subconjuntos de S.*

Novamente, temos uma nova terminologia para "mundo possível". Ao invés desse termo, utilizamos "situações possíveis". Porém, aqui não temos apenas uma mudança de termos, como foi o caso das alternativas epistêmicas. Essas últimas funcionam em Levesque como "situações bem comportadas"; os detalhes virão no que se segue.

Na lógica epistêmica clássica, temos apenas dois valores de verdade. Ou melhor, cada alternativa epistêmica sempre comporta ou a verdade ou a falsidade das fórmulas da linguagem. Nesse sentido, não existe uma proposição que seja e não seja o caso, ao mesmo tempo, em uma alternativa epistêmica; a não ser que entendamos as alternativas epistêmicas como os "impossíveis mundos possíveis" de Hintikka (1975), ou os mundos de domínios não-fixos de Rantala (1975).

[35] O termo "fórmula proposicional" significa que a fórmula em questão não contém B ou L.

De qualquer modo, as “situações” de Levesque ainda são diferentes. Na abordagem de Hintikka, para cada fórmula atômica p, temos que p deve ou não ser o caso em qualquer alternativa epistêmica. Ou seja, se w é uma alternativa epistêmica, deve haver $w \Vdash p$ ou $w \Vdash \neg p$, para cada w. Tendo isto, as alternativas epistêmicas de Hintikka funcionam em Levesque como situações logicamente completas (ou somente “situações completas”, se preferir)[36].

A ideia das “situações” na lógica de Levesque pode ser resumida no que se segue. Não é o caso que uma proposição primitiva[37] qualquer p é verdadeira ou falsa; ela pode ser verdadeira, falsa, ambas ou nenhuma. Basicamente, temos três tipos de situações possíveis:

1. Situações parciais: são situações que não suportam nem a verdade nem a falsidade de uma proposição primitiva p qualquer: isto é, $s \notin T(p) \cup F(p)$, sendo s uma situação.

2. Situações incoerentes: podemos, nessa lógica, ter situações incoerentes tais que suportam, ambas, a verdade e a falsidade de uma proposição p qualquer; isto é, $s \in T(p) \cap F(p)$. Essas situações, como já falamos antes, podem ser compreendidas como os “impossíveis mundos possíveis” de Hintikka.

3. Situações completas: suportam a verdade ou a falsidade de uma proposição p qualquer (mas não ambas). As situações completas não são incoerentes. Ou seja, elas são o mesmo que os “mundos possíveis” da semântica padrão da lógica epistêmica.

Uma definição de grande importância é a de “compatibilidade de situações”. Essa definição ajudará na checagem de validade de fórmulas:

Definição 2.10. *Compatibilidade de situações: uma situação completa s é compatível com uma situação s' se s e s' concordam sempre que s' é definida. Sendo assim, se $s' \in T(p)$, então $s \in T(p)$, e se $s' \in F(p)$, $s \in F(p)$, para cada proposição primitiva p.*

Para a checagem de validade, Levesque também define o conjunto $\mathscr{B}^*$, que consiste no conjunto de todas as situações completas em S compatíveis com alguma situação em $\mathscr{B}$.

[36]A explicação sobre os diferentes tipos de situação se encontra um pouco abaixo.

[37]*“Primitive proposition”*. Termo original utilizado por Levesque. Esse termo parece um tanto estranho; talvez devêssemos utilizar, no lugar dele, o termo “fórmula atômica”. De qualquer modo, optamos por manter o termo original.

Já falamos acerca das situações parciais, incoerentes e completas. Dissemos, por exemplo, que uma situação incoerente "suporta", ambas, a verdade e a falsidade de uma proposição primitiva p qualquer. Precisamos definir, então, as relações de suporte. Primeiramente, a notação e seu significado são os seguintes:

- Escrevemos $\eta, s \models_T \varphi$ quando queremos dizer que a situação s da estrutura η suporta a verdade da fórmula φ;
- Escrevemos $\eta, s \models_F \varphi$ quando queremos dizer que uma situação s da estrutura η suporta a falsidade da fórmula φ.

Em seguida, já podemos definir as relações de suporte para os operadores.

Definição 2.11. *As relações de suporte para os operadores da lógica das crenças implícitas e explícitas são:*

1. $\eta, s \models_T p$, *sendo* p *uma proposição primitiva, se, e somente se,* $s \in T(p)$;
2. $\eta, s \models_F p$, *sendo* p *uma proposição primitiva, se, e somente se,* $s \in F(p)$;
3. $\eta, s \models_T \neg\varphi$ *se, e somente se,* $\eta, s \models_F \varphi$, *e* $\eta, s \models_F \neg\varphi$ *se, e somente se,* $\eta, s \models_T \varphi$;
4. $\eta, s \models_T \varphi \wedge \psi$ *se, e somente se,* $\eta, s \models_T \varphi$ *e* $\eta, s \models_T \psi$;
5. $\eta, s \models_F \varphi \wedge \psi$ *se, e somente se,* $\eta, s \models_F \varphi$ *ou* $\eta, s \models_F \psi$;
6. $\eta, s \models_T B\varphi$ *se, e somente se,* $\eta, t \models_T \varphi$, *para todo* $t \in \mathscr{B}$;
7. $\eta, s \models_F B\varphi$ *se, e somente se,* $\eta, s \not\models_T B\varphi$;
8. $\eta, s \models_T L\varphi$ *se, e somente se,* $\eta, t \models_T \varphi$, *para todo* $t \in \mathscr{B}^*$;
9. $\eta, s \models_F L\varphi$ *se, e somente se,* $\eta, s \not\models_T L\varphi$.

Acompanhando a definição da relação de suporte para os operadores, temos outras definições igualmente importantes.

Definição 2.12. *Verdade ou satisfação de uma fórmula: uma fórmula* φ *é verdadeira (ou satisfeita) em uma situação* s*, se* $\eta, s \models_T \varphi$ *vale.*

Definição 2.13. *Validade: Uma fórmula* φ *é válida – simbolizamos como* $\models \varphi$ *– se* φ *é verdadeira em* s*, para todas as estruturas* $\eta = (S, \mathscr{B}, T, F)$*, e todas as situações completas* $s \in S$.

Antes de verificarmos os resultados da lógica de Levesque, convém deixarmos claro que, aqui, continuamos entendendo por onisciência lógica aquilo que é expresso pelo princípio (E-CLOS 2)[38]. Continuaremos, também, mantendo a distinção entre:

1. Fecho sob implicação (E-CLOS 1); isto é, se um agente acredita em φ e $\varphi \rightarrow \psi$, então o agente também acredita em ψ;

2. Fecho sob implicação válida (E-CLOS 6): se $\varphi \rightarrow \psi$ é válida e se o agente acredita em φ, então o agente também acredita em ψ;

3. Crenças (conhecimento) de fórmulas válidas (E-CLOS 3): se φ é uma fórmula válida, então o agente acredita em φ.

Alguns resultados importantes em Levesque

Com base no que foi definido até então, podemos observar algumas propriedades interessantes da lógica de Levesque. Em primeiro lugar, podemos verificar:

Teorema 2.13. $\models Bp \rightarrow Lp$ *– as crenças explícitas implicam as crenças implícitas.*

Demonstração. Suponha uma situação s qualquer de uma estrutura η qualquer tal que $\eta, s \models_T Bp$. Assim, pela definição 2.11-6, temos que $\eta, t \models_T p$, para todo $t \in \mathscr{B}$. Então, $t \in T(p)$, para todo $t \in \mathscr{B}$ (definição 2.11-1). Se $\mathscr{B}^*$ é o conjunto de todas as situações em S compatíveis com toda situação $t \in \mathscr{B}$, temos que, para todo $t \in \mathscr{B}^*$, $t \in T(p)$ (2.10). Deste modo, $\eta, t \models_T p$, para todo $t \in \mathscr{B}^*$ (definição 2.11-1). Logo, $\eta, s \models_T Lp$ (definição 2.11-8). Como ambas s e η foram escolhidas arbitrariamente, concluímos com $\models Bp \rightarrow Lp$. $\square$

Podemos, também, como aponta Fagin & Halpern (1988, p. 46), verificar que as crenças explícitas não possuem quaisquer das quatro propriedades de onisciência lógica apresentadas no início deste capítulo. Assim, para as crenças explícitas, os princípios (E-CLOS 1), (E-CLOS 2), (E-CLOS 3) e (E-CLOS 6) são todos **inválidos na lógica de Levesque**. Na lógica de Levesque, a fórmula $(Bp \wedge B(p \rightarrow q)) \wedge \neg Bq$ é satisfatível.

Teorema 2.14. *A fórmula* $(Bp \wedge B(p \rightarrow q)) \wedge \neg Bq$ *é satisfatível.*

[38]Consultar seção 2.2.

Demonstração. Suponha que haja uma estrutura qualquer η e uma situação s qualquer tal que $\eta, s \models_T Bp$ e $\eta, s \models_T B(p \to q)$. Sendo assim, $\eta, t \models_T p$ e $\eta, t \models_T p \to q$, para todo $t \in \mathscr{B}$ (definição 2.11-6). Os $t's$ considerados podem consistir de situações incoerentes. Tendo isto, pode ser o caso que $\eta, t \models_F q$, para todo $t \in \mathscr{B}$, apesar de $\eta, t \models_T p$ e $\eta, t \models_T p \to q$. Pois, se cada t é uma situação incoerente, ocorre $\eta, t \models_T q$ e $\eta, t \models_F q$. Ora, se para todo $t \in \mathscr{B}$ ocorre $\eta, t \models_F q$, pela contrapositiva da definição 2.11-6 obtemos $\eta, s \models_F Bq$. Assim, pela definição 2.11-3, chegamos a $\eta, s \models_T \neg Bq$. E portanto, nesse caso, $(Bp \wedge B(p \to q)) \wedge \neg Bq$ é satisfatível. □

Para complementar o teorema anterior, mostraremos que o operador B não é fechado sob implicação válida. Isto é, apesar da fórmula $p \to (p \wedge (q \vee \neg q))$ ser válida, a fórmula $Bp \wedge \neg B(p \wedge (q \vee \neg q))$ é satisfatível.

Teorema 2.15. *A fórmula $Bp \wedge \neg B(p \wedge (q \vee \neg q))$ é satisfatível.*

Demonstração. Suponha novamente uma estrutura η qualquer e uma situação s qualquer tal que $\eta, s \models_T Bp$. Então, $\eta, t \models_T p$, para todo $t \in \mathscr{B}$ (definição 2.11-6). Ora, os $t's$ considerados podem consistir de situações incoerentes. Em situações incoerentes, é aceitável sustentar a falsidade de fórmulas válidas como, por exemplo, $q \vee \neg q$[39]. Ou seja, podemos ter $\eta, t \models_F q \vee \neg q$, para todo $t \in \mathscr{B}$. Ora, se $\eta, t \models_F q \vee \neg q$, então $\eta, t \models_F p \wedge (q \vee \neg q)$ (definição 2.11-5). Agora, pela contrapositiva da definição 2.11-6, segue-se $\eta, s \models_F B(p \wedge (q \vee \neg q))$. Dai segue-se então, pela definição 2.11-3, $\eta, s \models_T \neg B(p \wedge (q \vee \neg q))$. Conclui-se, portanto, $\eta, s \models_T Bp \wedge \neg B(p \wedge (q \vee \neg q))$ (definição 2.11-4); isto é, a fórmula $Bp \wedge \neg B(p \wedge (q \vee \neg q))$ é satisfatível. □

Pode-se perceber também que nem todas as fórmulas válidas são necessariamente acreditadas; ou seja, $\models \varphi \;/\; \models B\varphi$ não é uma regra válida desta lógica.

[39] Em t, temos $\eta, t \models_T q$ e $\eta, t \models_F q$. Aplique agora a definição 2.11-3 em $\eta, t \models_T q$. Daí, $\eta, t \models_F \neg q$. A fórmula $q \vee \neg q$ pode ser falsa, deste modo, porque ocorre $\eta, t \models_F q$ e $\eta, t \models_F \neg q$. Estamos pensando aqui em definições usuais para o operador $\vee$:

- $\eta, s \models_T \varphi \vee \neg\varphi$ se, e somente se, $\eta, s \models_T \varphi$ ou $\eta, s \models_T \neg\varphi$;
- $\eta, s \models_F \varphi \vee \neg\varphi$ se, e somente se, $\eta, s \models_F \varphi$ e $\eta, s \models_F \neg\varphi$.

Porém, deve ser observado que, se $\eta, t \models_F q$ e $\eta, t \models_F \neg q$, pela definição 2.11-3 obtemos $\eta, t \models_T \neg q$ e $\eta, t \models_T q$. E daí, pela definição da relação de suporte para a disjunção, obtemos $\eta, s \models_T q \vee \neg q$. Este é um resultado um tanto esquisito, pois escolhemos pela falsidade de $q \vee \neg q$ quando essa fórmula também é verdadeira na situação incoerente. De qualquer forma esse é um recurso permitido pela lógica de Levesque. As situações incoerentes sustentam a verdade e a falsidade de uma proposição primitiva qualquer. Nesse caso, sustentar a verdade e a falsidade de q leva a sustentar a verdade e a falsidade de $q \vee \neg q$. Daí, escolhemos qualquer um dos resultados de acordo com o nosso propósito nessa prova, que nesse caso é $\eta, t \models_F q \vee \neg q$.

Para isto, temos uma prova muito simples, basta observar que o operador B não é fechado sob implicação válida.

Fagin & Halpern apontam também (1988, p. 46) que, na lógica de Levesque, é perfeitamente aceitável para o agente possuir crenças contraditórias.

Teorema 2.16. *A fórmula $Bp \wedge B\neg p$ é satisfatível.*

Demonstração. Suponha uma estrutura η qualquer e uma situação s qualquer. Pode ocorrer que toda situação $t \in \mathscr{B}$ seja incoerente[40]. Se isso for o caso, então $\eta, t \models_T p$ e $\eta, t \models_F p$, para todo $t \in \mathscr{B}$. Aplicando a definição 2.11-3 em $\eta, t \models_F p$, obtemos $\eta, s \models_T \neg p$, para todo $t \in \mathscr{B}$. Daí, pela definição 2.11-6, obtém-se $\eta, s \models_T Bp$ e $\eta, s \models_T B\neg p$. Tendo isto, aplicamos a definição 2.11-4 e chegamos a $\eta, s \models_T Bp \wedge B\neg p$. Portanto, a fórmula $Bp \wedge B\neg p$ é satisfatível. □

Apesar de o operador B não possuir quaisquer das propriedades de onisciência lógica que explicitamos, o mesmo não ocorre com o operador L. Não é difícil perceber, através das definições dadas até então, que a fórmula $(Lp \wedge L(p \to q)) \to Lq$ é válida.

Teorema 2.17. $\models (Lp \wedge L(p \to q)) \to Lq$

Demonstração. Suponha uma estrutura qualquer η e uma situação qualquer s tal que $\eta, s \models_T Lp$ e $\eta, s \models_T L(p \to q)$. Pela definição 2.11-8, temos $\eta, t \models_T p$ e $\eta, t \models_T p \to q$, para todo $t \in B^*$[41]. Logo, $\eta, t \models_T q$, para todo $t \in B^*$[42]. Sendo assim, pela definição 2.11-8, obtemos $\eta, s \models_T Lq$. Se, a partir das hipóteses $\eta, s \models_T Lp$ e $\eta, s \models_T L(p \to q)$ obtivemos $\eta, t \models_T Lq$, concluímos que a fórmula $(Lp \wedge L(p \to q)) \to Lq$ é válida, dado que tanto η quanto s foram escolhidos arbitrariamente. Portanto, na lógica de Levesque, as crenças implícitas são fechadas sob implicação, ao contrário das crenças explícitas. □

Outra propriedade dessa lógica é a seguinte: a regra $\models \varphi \;/\models L\varphi$ é válida. Ou seja, todos os agentes acreditam implicitamente em todas as fórmulas válidas.

[40]Intuitivamente, isso significa dizer que na dada situação s, todas as situações consideradas pelo agente como sendo alternativas podem ser incoerentes.

[41]Deve-se ter sempre em mente que B^* diz respeito apenas às situações completas.

[42]Nessa lógica, não há alterações na interpretação dos conectivos clássicos. Deste modo, as relações de suporte para $\to$ podem ser definidas de modo usual:

- $\eta, s \models_T \varphi \to \psi$ se, e somente se, $\eta, s \models_F \varphi$ ou $\eta, s \models_T \psi$;
- $\eta, s \models_F \varphi \to \psi$ se, e somente se, $\eta, s \models_T \varphi$ e $\eta, s \models_F \psi$.

Teorema 2.18. *Se* $\vDash \varphi$*, então* $\vDash L\varphi$*.*

Demonstração. Suponha uma fórmula válida qualquer; isto é, suponha que $\vDash \varphi$. Suponha então uma estrutura η qualquer, uma situação $s \in S$ qualquer, e também alguma situação $t \in S$ que seja compatível com s. Se $\vDash \varphi$, então $\eta, t \vDash_T \varphi$, para todo $t \in \mathscr{B}^*$ (pela definição 2.13). Daí, $\eta, s \vDash_T L\varphi$ (definição 2.11-8). Dado que tanto η quanto s foram escolhidos arbitrariamente, concluímos então com: Se $\vDash \varphi$, então $\vDash L\varphi$. □

De tudo o que foi colocado até então, podemos observar dois detalhes relevantes. O primeiro é $\nvDash (Bp \wedge B(p \rightarrow q)) \rightarrow Bp$. Ou seja, as crenças explícitas não são fechadas sob implicação. Com relação a este detalhe, o que realmente nos interessa no momento é aquilo que foi utilizado para a demonstração de seu contra-exemplo, isto é, $(Bp \wedge B(p \rightarrow q)) \wedge \neg Bq$. Na demonstração utilizamos as situações incoerentes.

Note-se, também, que a equivalência $B(p \wedge \neg p) \leftrightarrow (Bp \wedge B\neg p)$ é válida.

Teorema 2.19. $\vDash B(p \wedge \neg p) \leftrightarrow (Bp \wedge B\neg p)$

Demonstração. Para provar isso, precisamos provar duas coisas:

1. $\vDash B(p \wedge \neg p) \rightarrow (Bp \wedge B\neg p)$;

2. $\vDash (Bp \wedge B\neg p) \rightarrow B(p \wedge \neg p)$.

- Caso 1. $\vDash B(p \wedge \neg p) \rightarrow (Bp \wedge B\neg p)$

  Demonstração. Suponha $\eta, s \vDash_T B(p \wedge \neg p)$, para uma estrutura η e uma situação s qualquer. Assim, pela definição 2.11-6, $\eta, t \vDash_T p \wedge \neg p$, para todo $t \in \mathscr{B}$. Deste modo, $\eta, t \vDash_T p$ e $\eta, t \vDash_T \neg p$, para todo $t \in \mathscr{B}$ (definição 2.11-4). Daí, por 2.11-6, $\eta, s \vDash_T Bp$ e $\eta, s \vDash_T B\neg p$. E por 2.11-4, obtemos $\eta, s \vDash_T Bp \wedge \neg Bp$. Como $\eta, s \vDash_T B(p \wedge \neg p)$ é nossa hipótese inicial, concluímos $\eta, s \vDash_T B(p \wedge \neg p) \rightarrow (Bp \wedge B\neg p)$. Dado que tanto η quanto s foram escolhidos arbitrariamente, generalizamos o resultado para $\vDash B(p \wedge \neg p) \rightarrow (Bp \wedge B\neg p)$. □

- Caso 2. $\vDash (Bp \wedge B\neg p) \rightarrow B(p \wedge \neg p)$

Demonstração. suponha $\eta, s \models_T Bp \wedge B\neg p$, para uma estrutura η qualquer e uma situação s qualquer. Assim, por 2.11-4, $\eta, s \models_T Bp$ e $\eta, s \models_T B\neg p$. Deste modo, por 2.11-6, $\eta, t \models_T p$ e $\eta, t \models_T \neg p$, para todo $t \in \mathscr{B}$. Daí, $\eta, t \models_T p \wedge \neg p$, para todo $t \in \mathscr{B}$ (definição 2.11-4). E por 2.11-6, $\eta, s \models_T B(p \wedge \neg p)$. Como $\eta, s \models_T Bp \wedge B\neg p$ é nossa hipótese inicial, concluímos $\eta, s \models_T (Bp \wedge B\neg p) \rightarrow B(p \wedge \neg p)$. Dado que tanto η quanto s foram escolhidos arbitrariamente, generalizamos o resultado com $\models (Bp \wedge B\neg p) \rightarrow B(p \wedge \neg p)$. □

Tendo provado os casos 1 e 2, concluímos a prova do teorema. Portanto, $\models B(p \wedge \neg p) \leftrightarrow (Bp \wedge B\neg p)$. □

Em Levesque, o seguinte resultado também pode ser derivado:

- $(Bp \wedge B(p \rightarrow q)) \rightarrow B(q \vee (p \wedge \neg p))$

O esquema acima mostra que, ou o agente é logicamente onisciente, ou então há alguma situação que ele considera possível que é incoerente – isto é, o agente sofre de atenção desconexa.

Teorema 2.20. $\models (Bp \wedge B(p \rightarrow q)) \rightarrow B(q \vee (p \wedge \neg p))$

Demonstração. Suponha $\eta, s \models_T Bp$ e $\eta, s \models_T B(p \rightarrow q)$, para uma estrutura η e uma situação s qualquer. Assim, pela definição 2.11-6, $\eta, t \models_T p$ e $\eta, t \models_T p \rightarrow q$, para todo $t \in \mathscr{B}$. Temos a partir daí dois casos possíveis.

- <u>Caso 1</u>. Se $\eta, t \models_T p$ e $\eta, t \models_T p \rightarrow q$, então deve ocorrer $\eta, t \models_T q$, para todo $t \in \mathscr{B}$. Pois, $\eta, t \models_T p \rightarrow q$ se, e somente se, $\eta, t \models_F p$ ou $\eta, t \models_T q$. Como já se aceitou $\eta, t \models_T p$, devemos inferir $\eta, t \models_T q$.

- <u>Caso 2</u>. Nesse caso, assumimos o contrário; isto é, assumimos que, apesar de $\eta, t \models_T p$ e $\eta, t \models_T p \rightarrow q$, ocorre $\eta, t \models_F q$. Ora, se $\eta, t \models_T p \rightarrow q$, então $\eta, t \models_F p$ ou $\eta, t \models_T q$. Mas, foi assumido que $\eta, t \models_F q$. Deste modo, devemos ter $\eta, t \models_F p$. Daí, pela definição 2.11-3, $\eta, t \models_T \neg p$. Ora, já havíamos antes obtido $\eta, t \models_T p$, para todo $t \in \mathscr{B}$. Segue-se daí, por 2.11-4, $\eta, t \models_T p \wedge \neg p$, para todo $t \in \mathscr{B}$

Tendo isto, para todo $t \in \mathscr{B}$, ou $\eta, t \models_T q$ ou $\eta, t \models_T p \wedge \neg p$; isto é, $\eta, t \models_T q \vee (p \vee \neg p)$, para todo $t \in \mathscr{B}$. Sendo assim, pela definição 2.11-6, $\eta, s \models_T B(q \vee (p \wedge \neg p))$. Como hipótese inicial temos $\eta, s \vDash_T (Bp \wedge B(p \rightarrow q))$. Logo, $\eta, s \vDash_T (Bp \wedge B(p \rightarrow q)) \rightarrow B(q \vee (p \wedge \neg p))$. Dado que tanto η quanto s foram escolhidas arbitrariamente, generalizamos o resultado para $\vDash (Bp \wedge B(p \rightarrow q)) \rightarrow B(q \vee (p \wedge \neg p))$. $\square$

A partir deste resultado e do resultado $\vDash (Bp \wedge B\neg p) \leftrightarrow B(p \wedge \neg p)$, podemos levantar objeções contra a solução de Levesque. Observem o esquema $(Bp \wedge B\neg p) \leftrightarrow B(p \wedge \neg p)$. Ele diz que o agente pode ter crenças inconsistentes se, e somente se, toda situação que o agente considera possível for incoerente. Porém, como aponta Huang & Kwast, (1991, p. 7), "imaginar um agente que considera possível situações incoerentes é geralmente contra nossas intuições". Outra crítica levantada pelos autores é que a lógica de Levesque sofre de um problema crítico de representação, dado que a linguagem não permite reiteração das modalidades B e L (HUANG & KWAST, 1991, p. 7).

Apesar de havermos mostrado a atuação das situações incoerentes na falha de onisciência lógica, bem como na possibilidade de o agente ter crenças inconsistentes, também é possível adicionar[43] uma propriedade que considera a falha em onisciência lógica a partir da falha em "estar ciente" dos conceitos relevantes. Essa propriedade é inserida a partir da adição, à linguagem inicial, do operador A (de *awareness*). Em Levesque, o operador A pode ser também um dos motivos para a falha da onisciência lógica. Sendo assim, não é necessariamente por conta das situações incoerentes que o agente falha em ser logicamente onisciente, mas também por falta de "consciência" acerca de alguma proposição que é relevante para a derivação da crença.

O operador A funciona da seguinte maneira: $A\varphi$ é uma abreviação para 'o agente está ciente de φ'. A condição de suporte para A é: $\eta, s \vDash_T A\varphi$ se, e somente se, $\eta, s \vDash_T B(\varphi \vee \neg\varphi)$. Assim, uma situação s suporta a verdade de $A\varphi$ se, e somente se, suporta a verdade ou a falsidade de φ.

Com base em todas as definições apresentadas até então – incluindo o operador A – é possível notar o seguinte. Como já foi visto, nem todas as fórmulas válidas devem ser acreditadas explicitamente. Contudo, uma fórmula válida é acreditada tão logo um agente esteja ciente de todas as proposições primitivas que ocorrem nela. Deixando a ideia mais clara, temos o seguinte. Suponha por exemplo uma fórmula qualquer φ. Seja $Prim(\varphi)$[44] o conjunto das proposições primitivas

[43] Como mostraram FAGIN & HALPERN (1988, p. 47).
[44] Ver FAGIN & HALPERN, 1988, p. 47.

que ocorrem em φ. Seja $A\varphi$ a abreviação para Ap, sobre todo $p \in Prim(\varphi)$. Nessa interpretação, a seguinte proposição é satisfeita:

Teorema 2.21. *Se φ é uma fórmula proposicional válida, então* $\models A\varphi \rightarrow B\varphi$[45].

Considerações acerca da lógica de Levesque

Em geral, a lógica de Levesque é apropriada para captar a falha de onisciência lógica a partir da falha em estar ciente; além disso, também da falha em onisciência pela incapacidade de concentração em todos os temas simultaneamente – isto é, o que definimos por "atenção desconexa". Em se tratando de captar a falha de onisciência lógica por outros motivos, essa lógica não se mostra adequada. Esse fato nos leva então a duas opções. A primeira, nos leva à elaboração de uma lógica para crenças que seja capaz de captar um maior número de causas de falha em onisciência lógica. A outra alternativa é justamente discutir a relevância da primeira opção. Uma das maneiras de fazer isso é argumentar baseando-se em exemplos nos quais certos agentes falham em ser logicamente oniscientes em uma situação, mas são bem sucedidos em outras. Entraremos nessa discussão posteriormente.

A lógica de Levesque recebe uma série de críticas, algumas delas são de ordem filosófica, e outras simplesmente técnicas. Fagin & Halpern (1988, p. 48) destacam algumas dessas críticas:

1. A relação de suporte $\models_T$ é definida para todas as situações, sejam elas parciais, incoerentes ou completas. Todavia, ao observarmos a definição de validade, apenas as situações completas são consideradas quando se quer checar a validade de alguma fórmula. Isso vai contra a própria intuição dessa mesma lógica, já que a ideia é fazer uma semântica de situações possíveis, e não apenas de mundos possíveis bem comportados, como é o caso da semântica padrão de mundos possíveis proposta por Hintikka em *Knowledge and Belief*.

2. A intuição por trás de $\models$ não é clara quando estamos falando dos conectivos proposicionais. Para mostrar isso, suponha por exemplo um agente qualquer tal que $\eta, s \models_T \neg Ap$. Isto é, o agente não está ciente de uma proposição p qualquer. Sendo assim, pela definição do operador A, nem $\eta, s \models_T p$ nem $\eta, s \models_F p$ valem. Com isso, $\eta, s \models (p \leftrightarrow p)$ também não vale. Apesar desse resultado, é

[45]A prova desta proposição está presente no texto de Fagin & Halpern (1988, anexos).

perfeitamente possível imaginar um agente que não esteja ciente de uma proposição p, mas esteja ciente de algumas tautologias proposicionais, incluindo $p \leftrightarrow p$. Assim, $\eta, s \models p \leftrightarrow p$ valeria.

3. Já foi colocado antes (HUANG & KWAST, 1991, p. 7) que a lógica de Levesque também sofre de um sério problema de representação. Nessa lógica, as fórmulas são definidas de modo que não há reiteração dos operadores de crença. Assim, nem B nem L aparecem sob o escopo um do outro. Todavia, se argumenta que uma boa lógica epistêmica deveria capturar o raciocínio de um agente acerca de suas próprias crenças e convicções, e também o raciocínio desse mesmo agente acerca das crenças e convicções de outros agentes. Isso só seria possível, portanto, se houvesse reiteração das modalidades de crença (ou conhecimento).

4. De acordo com Vardi (1986, p. 294), a lógica das crenças explícitas e implícitas de Levesque "não faz o agente menos logicamente onisciente, mas apenas troca a lógica na qual o agente irá raciocinar". Dito isto, o agente pode não acreditar em todas as consequências lógicas de suas crenças, estando em uma lógica padrão, mas fazê-lo em uma lógica não-padrão. Vardi mostra, por exemplo, que os agentes da lógica de Levesque são logicamente oniscientes na lógica de Anderson & Belnap[46]. O problema é que não há qualquer motivo para pensar que um agente usa melhor suas capacidades racionais em uma lógica relevante do que em uma lógica clássica.

Discuti essas críticas em uma outra ocasião[47] e mostrei que, de certa forma, algumas delas já haviam sido antecipadas por Levesque. Ideia similar já foi defendida por Hadley; uma delas é com relação às situações incoerentes:

> Ao afirmar que as situações incoerentes podem ao menos ser imaginadas, apesar de não realizadas, Levesque parece estar sugerindo que nós construímos ao menos algumas situações como objetos *mentais* (ou internas), ao invés de porções dos mundos possíveis ou reais[48]. (HADLEY, 1987, p. 19)

---

[46]BELNAP, Alan R.; ANDERSON, Nuel D. **Entailment: the logic of relevance and necessity.** Princeton: Princeton University Press, 1975.

[47]MEDEIROS, Stanley K. B. **O problema da onisciência lógica e a lógica das crenças implícitas e explícitas de Levesque**. Comunicação proferida no V Encontro Interinstitucional de Filosofia - CCHLA - UFPB, João Pessoa, agosto de 2007.

[48]"*In claiming that incoherent situations can at least be imagined, though they cannot be realized, Levesque seems to be suggesting that we construe at least some situations as mental (or internal) objects, rather than portions of actual or possible worlds*."

2.3.2 Lógica da consciência

Além de abordagens como a de Levesque, há também as lógicas da consciência[49]. A seguir, discutiremos duas dessas lógicas, ressaltando algumas de suas propriedades mais importantes, bem como identificando e discutindo possíveis questões contra elas. A primeira delas podemos chamar simplesmente de "lógica da consciência". A outra, um pouco mais refinada do que a primeira, é chamada de "lógica da consciência geral". Vejamos então a primeira delas.

O foco central da lógica da consciência é a função A, de *awareness*. Esta função representa o estado de consciência ou, também, o ato de estar ciente de algo. A lógica da consciência é apresentada e discutida por Fagin & Halpern (1988, p. 49) como uma das alternativas de solução para o problema da onisciência lógica.

Podemos pensá-la, basicamente, como uma extensão da lógica de Levesque. Por ser justamente uma extensão, a lógica da consciência não só tem o poder representativo da lógica de Levesque, como também possui ainda algumas propriedades adicionais. Entre essas propriedades está o fato de ela não sofrer dos mesmos problemas de representação atribuídos à lógica das crenças implícitas e explícitas de Levesque. Na lógica da consciência, é permitida, por exemplo, a reiteração das modalidades. Isso nos dá uma riqueza imensa de material para discussão. Com essa capacidade representativa, é possível, por exemplo, discutir esquemas como $B_aB_b\varphi \rightarrow B_a\varphi$, $B_a\varphi \rightarrow B_aB_a\varphi$, ou $L_aL_b\varphi \rightarrow L_a\varphi$, $L_a\varphi \rightarrow L_aL_a\varphi$. O primeiro é conhecido como "transmissibilidade de conhecimento" (nesse caso, de crença); o segundo, é amplamente conhecido como "princípio da introspecção positiva". Essas noções nunca poderiam ser estudadas na lógica das crenças explícitas e implícitas de Levesque, já que esta última não permite fórmulas como $B_aB_b\varphi \rightarrow B_a\varphi$, $B_a\varphi \rightarrow B_aB_a\varphi$, ou $L_aL_b\varphi \rightarrow L_a\varphi$ e $L_a\varphi \rightarrow L_aL_a\varphi$.

Outra propriedade interessante da lógica da consciência é a permissão de agentes múltiplos, coisa que também é proibida em Levesque. É justamente essa propriedade também uma das responsáveis por permitir $B_aB_b\varphi \rightarrow B_a\varphi$ ou $L_aL_b\varphi \rightarrow L_a\varphi$ na linguagem, já que, observando, vemos claramente que esses esquemas são construídos com mais de um agente, além da reiteração de modalidades. Apesar das diferenças, a lógica da consciência ainda mantém muitas das propriedades da

[49]O termo original em inglês é *Logics of Awareness*. É possível que abusemos um pouco da linguagem, às vezes utilizando ambos "está ciente" ou "está consciente" para traduzir a noção *"to be aware of"*. Quando utilizarmos nesse contexto a palavra consciência, portanto, estamos pensando em *awareness*.

lógica de Levesque; algumas dessas propriedades veremos no decorrer desta seção.

Na lógica da consciência, novamente, há mais de um operador para a representação da crença:

Operador B: crenças explícitas.

Operador L: crenças implícitas.

Isto é, a distinção entre crenças implícitas e crenças explícitas é mantida. A linguagem é definida de modo usual. Como já foi explicitado, diferentemente da lógica de Levesque, é permitida a reiteração de modalidades. Os símbolos proposicionais primitivos também são $\wedge$ e $\neg$. Outros símbolos proposicionais, tais como $\rightarrow$, $\vee$ e $\leftrightarrow$ podem ser definidos posteriormente. Ao lidar com o caso de mais de um agente, nós temos $B_1,...,B_n$, $L_1,...,L_n$ operadores. É permitida, portanto, reiterações arbitrárias B_i e L_i em fórmulas. Pela similaridade das duas lógicas, não há necessidade de mostrar todas as cláusulas para construção de fórmulas. No entanto, achamos conveniente explicitar a cláusula que permite essas reiterações:

Se φ é uma fórmula, $B\varphi$ e $L\varphi$ também são fórmulas[50].

Na lógica da consciência, são dispensadas as situações parciais e incoerentes. A propósito, em lugar do termo "situação", usamos "estados". A noção é praticamente a mesma de "mundo possível". Um "estado" s é pensado como "estado de coisas". A ideia ficará mais clara quando descrevermos a semântica.

Definição 2.14. *Estrutura: Uma estrutura é uma quádrupla $\eta = (S, \pi, A = \{A_1,...,A_m\}, B = \{\mathscr{B}_1,...,\mathscr{B}_n\})$. Os elementos de η:*

- *S: conjunto de estados;*
- *π: atribuição de valor de verdade às sentenças primitivas para cada estado $s \in S$;*
- *$\mathscr{B}$: relação sobre os membros de S, para cada $i = 1,...,n$. $\mathscr{B}$ é uma relação serial, transitiva e euclideana;*

[50]Neste momento, façamos uma comparação com a lógica de Levesque. Ali, o termo "fórmula proposicional" aparecia para deixar claro que a fórmula φ não continha B ou L. Na cláusula atual, o termo não aparece. Isso significa que, sendo φ uma fórmula (não importando se já contém ou não B ou L), $B\varphi$ e $L\varphi$ também são fórmulas.

- *A: uma função de consciência A_i associa cada estado s a um conjunto de proposições primitivas. $A_i(s)$ consiste no conjunto de sentenças primitivas acerca das quais o agente i está ciente no estado s.*

Há também outra inovação com relação à lógica de Levesque. As relações de suporte $\models_T^\Psi$ e $\models_F^\Psi$ servem, digamos, para substituir as situações parciais. Os autores Fagin & Halpern (1988, p. 49) apontam o seguinte:

> Note-se que um estado corresponde a uma situação completa ou mundo possível. Não há estados parciais. Contudo, conseguimos alguns dos efeitos das situações parciais ao definir as relações de suporte $\models_T^\Psi$ e $\models_F^\Psi$, relativas a cada conjunto Ψ das proposições primitivas. Intuitivamente, o efeito de $\models_T^\Psi$ e $\models_F^\Psi$ é restringir todo estado a uma situação parcial na qual somente as proposições primitivas em Ψ são definidas[51].

A questão é: como analisar semanticamente a fórmula B_ip, considerando a função A_i e as relações de suporte $\models_T^\Psi$ e $\models_F^\Psi$? A ideia de Fagin & Halpern é que um "estado s suporta a verdade de $B_i\varphi$ relativo a Ψ somente se todos os estados que o agente i considera possíveis em s suportam a verdade de φ relativamente a $\Psi \cap A_i(s)$" (FAGIN & HALPERN, 1988, p. 49). Ou seja, Ψ é um conjunto específico de proposições primitivas das quais o agente i está ciente em s.

Note-se que, nesse caso, a função A é decisiva para dar valor de verdade a essa fórmula. O conjunto Ψ representa proposições primitivas importantes no contexto. São fatores de consciência que o agente requer para ter a crença em φ. Ou seja, o agente i precisa estar consciente de certas proposições primitivas para poder formular uma crença em φ; ou ainda, a crença em φ requer para i a consciência acerca de certos fatores decisivos a sua obtenção.

Além das relações de suporte $\models_T^\Psi$ e $\models_F^\Psi$, há também uma relação de verdade padrão $\models$. Diz-se que uma fórmula $B_i\varphi$ é verdadeira em um estado s, se s suporta a verdade de $B_i\varphi$ relativamente a $A_i(s)$.

Apresentemos formalmente as relações de suporte.

Definição 2.15. *As relações de suporte para a lógica da consciência são:*

---

[51] *"Note that a state corresponds to a complete situation or possible world. There are no partial states. However, we get some of the effects of taking partial states by defining support relations $\models_T^\Psi$ and $\models_F^\Psi$ relative to each set Ψ of primitive propositions. Intuitively, the effect of $\models_T^\Psi$ and $\models_F^\Psi$ is to restrict every state to a partial situation where only the primitive propositions in Ψ are defined."*

1. $\eta, s \models_T^\Psi true$[52];

2. $\eta, s \not\models_F^\Psi true$;

3. $\eta, s \models true$;

4. $\eta, s \models_T^\Psi \varphi$, *sendo* φ *uma proposição primitiva, se e somente se* $\pi(s, \varphi) = V$ *e* $\varphi \in \Psi$;

5. $\eta, s \models_F^\Psi \varphi$, *sendo* φ *uma proposição primitiva, se, e somente se,* $\pi(s, \varphi) = F$ *e* $\varphi \in \Psi$;

6. $\eta, s \models \varphi$, *sendo* φ *uma proposição primitiva, se e somente se* $\pi(s, \varphi) = V$;

7. $\eta, s \models_T^\Psi \neg\varphi$ *se, e somente se,* $\eta, s \models_F^\Psi \varphi$;

8. $\eta, s \models_F^\Psi \neg\varphi$ *se, e somente se,* $\eta, e \models_T^\Psi \varphi$;

9. $\eta, s \models \neg\varphi$ *se, e somente se,* $\eta, s \not\models \varphi$;

10. $\eta, s \models_T^\Psi \varphi \wedge \psi$ *se, e somente se,* $\eta, s \models_T^\Psi \varphi$ *e* $\eta, s \models_T^\Psi \psi$;

11. $\eta, s \models_F^\Psi \varphi \wedge \psi$ *se e somente se* $\eta, s \models_F^\Psi \varphi$ *ou* $\eta, s \models_F^\Psi \psi$;

12. $\eta, s \models \varphi \wedge \psi$ *se, e somente se,* $\eta, s \models \varphi$ *e* $\eta, s \models \psi$;

13. $\eta, s \models_T^\Psi B_i\varphi$ *se, e somente se,* $\eta, t \models_T^{\Psi \cap A_i(s)} \varphi$ *para todo* t *tal que* $(s,t) \in \mathscr{B}_i$;

14. $\eta, s \models_F^\Psi B_i\varphi$ *se, e somente se,* $\eta, t \models_F^{\Psi \cap A_i(s)} \varphi$, *para algum* φ *tal que* $(s,t) \in \mathscr{B}_i$;

15. $\eta, s \models B_i\varphi$ *se, e somente se,* $\eta, s \models_T^\Phi \varphi$, *sendo* Φ *o conjunto de todas as proposições primitivas;*

16. $\eta, s \models_T^\Psi L_i\varphi$ *se, e somente se,* $\eta, t \models_T^\Psi \varphi$ *para todo* t *tal que* $(s,t) \in \mathscr{B}_i$;

17. $\eta, s \models_F^\Psi L_i\varphi$ *se, e somente se,* $\eta, t \models_F^\Psi \varphi$, *para algum* t *tal que* $(s,t) \in \mathscr{B}_i$;

18. $\eta, s \models L_i\varphi$ *se e somente s se* $\eta, t \models \varphi$ *para todo* t *tal que* $(s,t) \in \mathscr{B}_i$.

Com todas as condições de suporte estabelecidas, fica restando a definição de validade. Com ela, parte-se para os resultados. A definição de validade apresentada por Fagin & Halpern (1988. p. 50) é basicamente a mesma da lógica de Levesque:

Definição 2.16. *Validade: uma fórmula* φ *é válida se* $\eta, s \models \varphi$, *para todas as estruturas* η *e todos os estados* s *em* η.

[52] Ao contrário da lógica de Levesque, aqui a fórmula *true* é adotada; neste caso, como constante de verdade.

Alguns resultados na lógica da consciência

Entre os resultados apontados por eles, está:

Teorema 2.22. *Na lógica da consciência, valem as seguintes propriedades:*

1. *A relação de suporte $\models$ é completa. Isto é, para cada η, s, ou $\eta, s \models \varphi$ ou $\eta, s \models \neg\varphi$.*

2. *(a) Se $\Psi \subseteq \Psi'$ e $\eta, s \models_T^{\Psi} \varphi$, então $\eta, s \models_T^{\Psi'} \varphi$.*
   (b)Se $\Psi \subseteq \Psi'$ e $\eta, s \models_F^{\Psi} \varphi$, então $\eta, s \models_F^{\Psi'} \varphi$.

3. *(a) Para cada conjunto Ψ de proposições primitivas, se $\eta, s \models_T^{\Psi} \varphi$ então $\eta, s \models \varphi$.*
   (b) Para cada conjunto Ψ de proposições primitivas, se $\eta, s \models_F^{\Psi} \varphi$ então $\eta, s \models \neg\varphi$.

4. *$\models B_i\varphi \to L_i\varphi$.*

Façamos a prova do primeiro e do último caso, isto é, itens 1 e 4.

Demonstração. A definição 2.15-9 garante que sempre é o caso que $\eta, s \models \neg\varphi$ se, e somente se, $\eta, s \not\models \varphi$. Assim, supondo $\eta, s \models \neg\varphi$ chegamos a $\eta, s \not\models \varphi$, e vice-versa. Nunca ocorre deste modo $\eta, s \models \neg\varphi$ e $\eta, s \models \varphi$ ao mesmo tempo. Portanto, $\eta, s \models \varphi$ ou $\eta, s \models \neg\varphi$ (sendo a disjunção exclusiva). □

O item quatro é exatamente a mesma propriedade da Lógica de Levesque: as crenças explícitas implicam as crenças implícitas.

Demonstração. Suponha uma estrutura η qualquer e um estado s qualquer dessa mesma estrutura. Suponha também que $\eta, s \models B_i\varphi$. Assim, pela definição 2.15-15, temos $\eta, s \models_T^{\Phi} B_i\varphi$, sendo Φ o conjunto de todas as proposições primitivas. Ora, se $\eta, s \models_T^{\Phi} B_i\varphi$, em particular, temos também $\eta, s \models_T^{\Psi} B_i\varphi$. Deste modo, pela definição 2.15-13, $\eta, t \models_T^{\Psi \cap A(s)} \varphi$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Pela hipótese $\eta, s \models B_i\varphi$ e pela definição 2.15-15, garante-se que $\eta, t \models_T^{\Psi} \varphi$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Pelo item 3.(a) do teorema 2.22, obtém-se $\eta, t \models \varphi$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Deste modo, por 2.15-18, $\eta, s \models L_i\varphi$. Logo, como η e s foram escolhidos arbitrariamente, segue-se que $\models B_i\varphi \to L_i\varphi$. □

Nesta lógica da consciência, assim como em Levesque, os agentes também acreditam implicitamente em todas as fórmulas válidas.

Teorema 2.23. *Se* $\vDash \varphi$*, então* $\vDash L_i\varphi$*.*

Demonstração. Suponha uma estrutura η qualquer e um estado s qualquer dessa mesma estrutura. Suponha então uma fórmula φ qualquer tal que $\vDash \varphi$. Como a relação $\mathscr{B}_i$ é serial, o estado s possui ao menos um estado t tal que $(s,t) \in \mathscr{B}_i$. Se $\vDash \varphi$, pela definição 2.16 obtemos $\eta, t \vDash \varphi$. Como t foi escolhido arbitrariamente, podemos dizer que $\eta, t \vDash \varphi$ vale para todos os t´s de η, $(s,t) \in \mathscr{B}_i$. Deste modo, pela definição 2.15-18, $\eta, s \vDash L_i\varphi$. Logo, como η e s também foram escolhidos arbitrariamente, podemos generalizar o resultado e obter: Se $\vDash \varphi$, então $\vDash L_i\varphi$. □

Considerando as outras duas propriedades de onisciência lógica, verifica-se também que as crenças implícitas são fechadas sob implicação, e também sob implicação válida. Vejamos primeiro o fecho sob implicação. Para a prova do teorema seguinte, acrescentamos duas cláusulas à definição 2.15:

19 $\eta, s \vDash_T^\Psi \varphi \to \psi$ se, e somente se, $\eta, s \vDash_F^\Psi p$ ou $\eta, s \vDash_T^\Psi q$;

20 $\eta, s \vDash_F^\Psi \varphi \to \psi$ se, e somente se, $\eta, s \vDash_T^\Psi p$ e $\eta, s \vDash_F^\Psi q$.

Teorema 2.24. $\vDash (L_i p \wedge L_i(p \to q)) \to L_i q$

Demonstração. Suponha uma estrutura η qualquer e um estado s qualquer da mesma estrutura. Suponha que $\eta, s \vDash_T^\Psi L_i p$ e $\eta, s \vDash_T^\Psi L_i(p \to q)$. Assim, pela definição 2.15-16, $\eta, t \vDash_T^\Psi p$ e $\eta, s \vDash_T^\Psi p \to q$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Deste modo, por 2.15-19, temos $\eta, t \vDash_T^\Psi q$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Portanto, pela definição 2.15-16, $\eta, s \vDash_T^\Psi L_i q$. Como η e s foram ambos escolhidos arbitrariamente, concluímos com $\vDash (L_i p \wedge L_i(p \to q)) \to L_i q$; ou seja, $(L_i p \wedge L_i(p \to q)) \to L_i q$ é uma fórmula válida da lógica da consciência. □

A prova do fecho sob implicação válida, por parte das crenças implícitas, se dá quase do mesmo modo. Para uma prova satisfatória da propriedade, é preciso mostrar o seguinte:

Teorema 2.25. *Se para um estado* s *qualquer de uma estrutura* η *qualquer,* $\eta, s \vDash_T^\Psi L_i p$ *e* $\vDash p \to q$*, então* $\eta, s \vDash_T^\Psi L_i q$*.*

Demonstração. Suponha então um estado s qualquer de uma estrutura η qualquer tal que $\eta, s \vDash_T^\Psi L_i p$. Suponha também $\vDash p \to q$. Assim, pela definição 2.16, $\eta, t \vDash_T^\Psi$

$p \to q$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Tendo isto, pela definição 2.15-18 obtemos $\eta, s \models_T^\Psi L_i(p \to q)$. Ora, se temos $\eta, s \models_T^\Psi L_i p$ e $\eta, s \models_T^\Psi L_i(p \to q)$, o teorema 2.24 nos garante $\eta, s \models_T^\Psi L_i q$. O estado s e estrutura η foram escolhidos arbitrariamente. Logo, se para um estado s qualquer de uma estrutura η qualquer, $\eta, s \models_T^\Psi L_i p$ e $\models p \to q$, então $\eta, s \models_T^\Psi L_i q$. □

Vale relembrar um detalhe importante da lógica da consciência: não há situações parciais ou incoerentes. Sendo assim, um estado sempre suporta a verdade de uma fórmula válida da lógica proposicional. Apesar disso, uma propriedade dessa lógica é que nem todas as fórmulas válidas são necessariamente acreditadas explicitamente. Isso vai acontecer justamente quando o agente não estiver ciente das proposições primitivas que constituem essa fórmula. Podemos, então, encontrar uma estrutura mental[53] em que $\neg B_i(p \vee \neg p)$ é satisfatível.

Teorema 2.26. *A fórmula* $\neg B_i(p \vee \neg p)$ *é satisfatível.*

Demonstração. Considere a fórmula proposicionalmente válida $p \vee \neg p$. Ou seja, $\models p \vee \neg p$. Pela definição 2.16, para qualquer estrutura η e estado s qualquer, $\eta, s \models_T^\Psi p \vee \neg p$. Porém, é possível que, em uma estrutura η e um estado t qualquer, $(s,t) \in \mathscr{B}_i$, tenhamos $\eta, t \models_F^{\Psi \cap A(s)} p \vee \neg p$. Isso é o caso porque pode ocorrer, na estrutura em questão, $p \notin A_i(s)$[54]. Dado que $\eta, t \models_F^{\Psi \cap A(s)} p \vee \neg p$, pela definição 2.15-14 obtém-se $\eta, s \models_F^\Psi B_i(p \vee \neg p)$. Daí, pela definição 2.15-7, obtém-se $\eta, s \models_T^\Psi \neg B_i(p \vee \neg p)$. Logo, $\neg B_i(p \vee \neg p)$ é satisfatível. □

Isso nos prova que nem sempre os agentes acreditam explicitamente em fórmulas válidas. Contudo, devemos observar que, nesse caso, a noção de consciência foi fundamental. O agente falhou em acreditar na fórmula válida $p \vee \neg p$ exatamente por não estar ciente acerca da proposição primitiva p. Se o agente por acaso estivesse ciente de p, ele consequentemente acreditaria explicitamente em $p \vee \neg p$, dado que todos os estados são completos.

Há também um resultado que é consonante com aquele encontrado na lógica das crenças implícitas e explícitas de Levesque: as crenças explícitas não são

[53]Termo utilizado por WHITSEY, Mark. **Logical omniscience: a survey**. Techical report NOTTCS-WP-2003-2. Nottinghan: University of Nottinghan, 2003.

[54]Intuitivamente, o agente pode não estar ciente acerca de p, sendo nesse caso relevante para a determinação da relação de suporte para fórmula $p \vee \neg p$. O ato de estar ciente ou não varia de estrutura mental para estrutura mental

fechadas sob implicação válida. Não é difícil entender o motivo. Acabamos de mostrar, no teorema 2.26, que nem sempre o agente acredita em uma fórmula válida. Todavia, para provar que as crenças explícitas não são fechadas sob implicação válida, é preciso algo mais; a falha do fecho sob implicação válida para as crenças explícitas é identificada ao se provar[55], no sistema, que a fórmula $B_i p \wedge \neg B_i(p \wedge (q \vee \neg q))$ é satisfatível (como em Levesque).

Teorema 2.27. *A fórmula $B_i p \wedge \neg B_i(p \wedge (q \vee \neg q))$ é satisfatível.*

Demonstração. Suponha $\vDash q \vee \neg q$ e uma estrutura mental qualquer η tal que $\eta, s \vDash_T^\Psi B_i p$. Pela definição 2.15-13, $\eta, t \vDash_T^{\Psi \cap A_i(s)} p$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Como foi provado anteriormente no teorema 2.26, nem sempre uma fórmula válida é acreditada explicitamente. Podemos supor então que η é uma estrutura e s é um estado tal que $\eta, s \vDash_F^\Psi B_i(q \vee \neg q)$, já que na estrutura $q \notin \Psi \cap A_i(s)$. Deste modo, pela definição 2.15-14, $\eta, t \vDash_F^{\Psi \cap A_i(s)} q \vee \neg q$, para algum t tal que $(s,t) \in \mathscr{B}_i$. Se isso é o caso, por 2.15-11, $\eta, t \vDash_F^{\Psi \cap A_i(s)} p \wedge (q \vee \neg q)$. Daí, pela definição 2.15-14, $\eta, s \vDash_F^\Psi B_i(p \wedge (q \vee \neg q))$. E agora, pela definição 2.15-7, $\eta, s \vDash_T^\Psi \neg B_i(p \wedge (q \vee \neg q))$. Inicialmente havíamos admitido $\eta, s \vDash_T^\Psi B_i p$. Logo, se temos $\eta, s \vDash_T^\Psi B_i p$ e $\eta, s \vDash_T^\Psi \neg B_i(p \wedge (q \vee \neg q))$, pela definição 2.15-10 obtemos $\eta, s \vDash_T^\Psi B_i p \wedge \neg B_i(p \wedge (p \vee \neg p))$. □

Vimos, então, que na lógica da consciência as crenças explícitas não são fechadas sob implicação válida. Porém, as crenças explícitas são fechadas sob implicação. É dizer que, se em uma mesma estrutura mental, o agente acredita explicitamente em uma fórmula p e na implicação $p \rightarrow q$, então o agente também acredita explicitamente em q. O que permite a lógica da consciência livrar-se do fecho sob implicação válida é o fato de a mesma admitir várias estruturas mentais. Deste modo, podemos considerar estruturas $\eta ... \eta_n$ nas quais o agente está ou não ciente de certas proposições primitivas. No caso da fórmula proposicionalmente válida $q \vee \neg q$, o agente falhou em acreditar explicitamente nela exatamente por não estar ciente da proposição primitiva q. Não podemos, portanto, negar a possibilidade de outra estrutura η' em que o agente está de fato ciente acerca da proposição primitiva q. Se isso for o caso, $\eta' \vDash_T^\Psi q \vee \neg q$. Devemos ter em mente que todos os estados da lógica da consciência são completos. Ao definirmos a implicação de modo usual e supor que, em um estado s de uma mesma estrutura η, $\eta, s \vDash_T^\Psi p$ e $\eta, s \vDash_T^\Psi p \rightarrow q$, temos que aceitar como consequência $\eta, s \vDash_T^\Psi q$. Para provar então que

[55]FAGIN & HALPERN, 1988, p. 50.

as crenças explícitas são fechadas sob implicação, basta mostrarmos que a fórmula $\eta, s \models_T^\Psi (B_i p \wedge B_i(p \rightarrow q)) \rightarrow B_i q$ é válida.

Teorema 2.28. $\models (B_i p \wedge B_i(p \rightarrow q)) \rightarrow B_i q$

Demonstração. Suponha $\eta, s \models_T^\Psi B_i p$ e $\eta, s \models_T^\Psi B_i(p \rightarrow q)$, para uma estrutura η e um estado s qualquer. Temos então, pela definição 2.15-13, $\eta, t \models_T^{\Psi \cap A_i(s)} p$ e $\eta, t \models_T^{\Psi \cap A_i(s)} p \rightarrow q$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Assim, por 2.15-19, obtém-se $\eta, t \models_T^{\Psi \cap A_i(s)} q$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Daí, novamente pela definição 2.15-13, obtém-se $\eta, s \models_T^\Psi B_i q$. Com base então em $\eta, s \models_T^\Psi B_i p$ e $\eta, s \models_T^\Psi B_i(p \rightarrow q)$ chegamos a $\eta, s \models_T^\Psi B_i q$. Logo, $\eta, s \models_T^\Psi (B_i p \wedge B_i(p \rightarrow q)) \rightarrow B_i q$. Como η e s foram escolhidos arbitrariamente, podemos concluir com o resultado $\models (B_i p \wedge B_i(p \rightarrow q)) \rightarrow B_i q$. □

Teorema 2.29. *A fórmula $B_i(p \wedge \neg p)$ não é satisfatível.*

Demonstração. Suponha, por exemplo, uma estrutura qualquer η e um estado s qualquer da mesma estrutura. Vimos no teorema 2.22 que $\models$ é completa. Sendo assim, ou $\eta, s \models p$ ou $\eta, s \models \neg p$, mas não ambos. Deste modo, não pode ocorrer $\eta, s \models p$ e $\eta, s \models \neg p$ ao mesmo tempo. Por conta disso, para todo estado s, $\eta, s \models_F^\Psi p \wedge \neg p$. E para todo t, $(s,t) \in \mathscr{B}$, $\eta, t \models_F^{\Psi \cap A_i(s)} p \wedge \neg p$. Segue daí, por 2.15-14, $\eta, s \models_F^\Psi B_i(p \wedge \neg p)$; isto é, não há como um agente formular uma crença em $p \wedge \neg p$. Portanto, a fórmula $B_i(p \wedge \neg p)$ não é satisfatível em qualquer estrutura (dado que η foi escolhida arbitrariamente). □

Fagin & Halpern argumentam (1988, p. 51) que, em sua lógica da consciência, as crenças implícitas satisfazem todos os axiomas e regras do sistema tradicionalmente conhecido como "S5 fraco"[56]:

Teorema 2.30. *As crenças implícitas satisfazem todos os axiomas e regras listadas abaixo:*

1. *A1: Todas as instâncias das tautologias proposicionais;*
2. *A2:* $(L_i p \wedge (L_i(p \rightarrow q)) \rightarrow L_i q$;
3. *A3:* $\neg L_i(false)$;
4. *A4:* $L_i p \rightarrow L_i L_i p$;

[56] Apresentado brevemente em Fagin & Halpern (1988, p. 43).

5. *A5:* $\neg L_i p \rightarrow L_i \neg L_i p$;

6. *R1: Se* p, $p \rightarrow q$, *então* q;

7. *R2: Se* p, *então* $L_i p$.

O axioma A1 segue quase imediatamente, dado que todo estado s da lógica da consciência é completo, e por isso se comporta como um mundo possível da lógica tradicional. Contudo, vejamos com detalhes.

Demonstração. Item 1. O que queremos mostrar é que, se φ é uma tautologia (ou seja, se φ é uma fórmula válida da lógica proposicional), então $\vDash \varphi$ (φ é uma fórmula válida da lógica da consciência)[57]. Suponha então uma tautologia qualquer φ. Se φ é uma tautologia, então φ é verdadeira em qualquer valoração. Cada estado da lógica da consciência pode ser entendido como uma espécie de valoração, suportando a verdade ou a falsidade de uma fórmula. Já foi dito que todo estado s da lógica da consciência é completo. Suponha então um estado s qualquer de uma estrutura η qualquer. Pela definição de tautologia temos que, independentemente de suportar a verdade ou falsidade dos constituintes atômicos de φ, $\eta, s \vDash_T^\Psi \varphi$. Ou seja, s suporta a verdade de φ, não importando os valores que tomam seus constituintes. Daí, como η e s são arbitrários, pela definição de validade temos que $\vDash \varphi$. □

Demonstração. Item 2. Já demonstrado (teorema 2.24). □

A fórmula $false$ é a abreviação da fórmula $\neg true$. Agora provaremos a validade de A3, isto é, provaremos que $\vDash \neg L_i(false)$. Dizer que $\neg L_i(false)$ é válido é o mesmo que dizer $\eta, s \vDash_T^\Psi \neg L_i(false)$, para qualquer estado s de qualquer estrutura η.

Demonstração. Item 3. Suponha então um estado s qualquer de uma estrutura η qualquer. A relação $\mathscr{B}$ foi definida como sendo serial. Sendo assim, para cada $s \in S$, há algum t tal que $(s,t) \in \mathscr{B}_i$. Pela definição 2.15-1, temos $\eta, t \vDash_T^\Psi true$. Assim, pela definição 2.15-8, $\eta, t \vDash_F^\Psi \neg true$. Já foi dito que $false$ é uma abreviação para $\neg true$. Daí, $\eta, t \vDash_F^\Psi false$. E pela definição 2.15-17, $\eta, s \vDash_F^\Psi L_i(false)$. Agora, pela definição 2.15-7, chegamos a $\eta, s \vDash_T^\Psi \neg L_i(false)$. Como ambos η e s foram escolhidos arbitrariamente, conluímos generalizando o resultado para $\vDash \neg L_i(false)$. □

[57]Entendemos "tautologia" aqui como em Chellas (1980, p. 8), isto é, como uma sentença verdadeira em qualquer valoração dada aos seus constituintes atômicos.

Ao observarmos com cuidado, não é difícil notar que a validade do axioma A3 é obtida pelo fato de a relação $\mathscr{B}$ ser serial. O axioma A3 traz consequências interessantes. Ele diz que os agentes não podem acreditar na falsidade. Vale ressaltar que isso é diferente de dizer que um agente não pode ter crenças falsas a respeito de algo. Na verdade, é absolutamente normal possuir crenças que mais tarde se mostram falsas. Acreditar diretamente na falsidade é portanto diferente de ter uma simples crença falsa.

O axioma A4 segue facilmente pelo fato de a relação $\mathscr{B}$ ser transitiva. Como, na lógica da consciência, a relação $\mathscr{B}$ foi definida como sendo transitiva, obtém-se $\models L_i\varphi \rightarrow L_iL_i\varphi$. A estratégia para provar esse teorema é praticamente a mesma da anterior. Precisamos mostrar que $\eta, s \models_T^\Psi L_i\varphi \rightarrow L_iL_i\varphi$, para qualquer estado s de qualquer estrutura η.

Demonstração. Item 4 Suponha uma estrutura qualquer η e um estado qualquer s tal que $\eta, s \models_T^\Psi L_i p$. Daí, pela definição 2.15-16, $\eta, t \models_T^\Psi p$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Sabemos que há ao menos um t' tal que $(t,t') \in \mathscr{B}_i$. Ora, se a relação B é transitiva, e temos $(s,t) \in \mathscr{B}_i$ e $(t,t') \in \mathscr{B}_i$, segue-se então, pela definição 2.15-16, $\eta, t' \models_T^\Psi p$ para todo t' tal que $(t,t') \in \mathscr{B}_i$. Deste modo, novamente pela definição 2.15-16, $\eta, t \models_T^\Psi L_i p$, para todo t tal que $(s,t) \in \mathscr{B}_i$. E ainda pela definição 2.15-16, $\eta, s \models_T^\Psi L_iL_i p$. Como η e s foram escolhidos arbitrariamente, concluímos com $\models L_i p \rightarrow L_iL_i p$. $\square$

O Axioma A4 é defendido por Hintikka em *Knowledge and Belief*, e muito discutido desde então. Na verdade, o princípio da introspecção positiva é um dos grandes problemas da epistemologia, particularmente relevante na discussão sobres as diferenças entre as perspectivas de primeira/terceira pessoas no ato de atribuição de conhecimento.

Um outro axioma de introspecção é o axioma A5. Assim como o axioma A4, ele diz que todo agente tem conhecimento pleno acerca do conjunto de todas as suas crenças. Vejamos como o axioma A5 para crenças implícitas é válido na lógica da consciência.

Demonstração. Item 5 Suponha um estado s qualquer de uma estrutura η qualquer tal que $\eta, s \models_T^\Psi \neg L_i p$. Tendo isto, aplicamos a definição 2.15-7 e obtemos $\eta, s \models_F^\Psi L_i p$. Se isto é o caso, aplicamos a definição 2.15-17 e obtemos que $\eta, u \models_F^\Psi p$, para algum u tal que $(s,u) \in \mathscr{B}_i$. Suponha agora, para redução ao absurdo, que $\eta, s \models_F^\Psi L_i \neg L_i p$. Deste modo, pela definição 2.15-17, $\eta, t \models_F^\Psi \neg L_i p$, para algum t tal que

$(s,t) \in \mathscr{B}_i$. Ora, se $(s,t) \in \mathscr{B}_i$ e $(s,u) \in \mathscr{B}_i$, e a relação $\mathscr{B}_i$ é euclideana, obtemos $(t,u) \in \mathscr{B}_i$. Agora, se $\eta, t \vDash_F^\Psi \neg L_i p$ é o caso, aplicamos a definição 2.15-8 e obtemos $\eta, t \vDash_T^\Psi L_i p$. Daí, se $(t,u) \in \mathscr{B}_i$, pela definição 2.15-16 segue-se que $\eta, u \vDash_T^\Psi p$. Ora, mas isso é uma contradição com o resultado obtido anteriormente, a saber, $\eta, u \vDash_F^\Psi p$. Obtemos então $\eta, s \vDash_T^\Psi L_i \neg L_i p$. E assim, $\eta, s \vDash_T^\Psi \neg L_i p \to L_i \neg L_i p$. Como ambos η e s foram escolhidos arbitrariamente, concluímos com o resultado $\vDash \neg L_i p \to L_i \neg L_i p$. □

Com relação às regras R1 e R2, a prova é parecida com a que fazemos em lógica modal alética. Devemos, portanto, provar que as referidas regras preservam validade. A estratégia é supor que os antecedentes são válidos, para então mostrar que os consequentes também o são.

Demonstração. Item 5. Temos de mostrar que, se $\vDash p$ e $\vDash p \to q$, então $\vDash q$. Suponha então $\vDash p$ e $\vDash p \to q$. Assim, pela definição 2.16, $\eta, s \vDash_T^\Psi p$ e $\eta, s \vDash_T^\Psi p \to q$, para qualquer estado s de qualquer estrutura η. Deste modo, por 2.15-19, $\eta, s \vDash_T^\Psi q$, para qualquer s de qualquer η. Daí, por 2.16, obtém-se $\vDash q$. □

Demonstração. Item 6. Suponha uma fórmula válida qualquer p. Pela definição 2.16, $\eta, s \vDash_T^\Psi p$, para todas as estruturas η e todos os estados s em η. Como todos os estados s possuem ao menos um estado t tal que $(s,t) \in \mathscr{B}_i$, podemos inferir que $\eta, s \vDash_T^\Psi L_i p$, pois certamente $\eta, t \vDash_T^\Psi p$, também pela definição 2.16. Logo, $\vDash L_i p$. □

Nossa preocupação agora se volta para certas propriedades que representam motivo de discussão na lógica de Levesque. Foi possível observar ali que não era permitido reiteração de modalidades. Igualmente foi dito que a lógica da consciência, por permitir essas reiterações, escapa a certas críticas contra a lógica de Levesque. A reiteração de modalidades nos permite, por exemplo, discutir esquemas relevantes como $B_i L_i p \leftrightarrow B_i p$. No caso da lógica da consciência, tal esquema é considerado válido.

Teorema 2.31. $\vDash B_i L_i p \leftrightarrow B_i p$.

Demonstração. Para provarmos essa propriedade, precisamos provar $\vDash B_i L_i p \to B_i p$ e $\vDash B_i p \to B_i L_i p$.

Demonstração. Caso 1: $\vDash B_i L_i p \to B_i p$.

Suponha um estado qualquer s de uma estrutura qualquer η tal que $\eta, s \models^{\Psi}_{T} B_i L_i p$. Deste modo, pela definição 2.15-13, $\eta, t \models^{\Psi \cap A_i(s)}_{T} L_i p$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Assim, pela definição 2.15-16, $\eta, u \models^{\Psi \cap A_i(s)}_{T} p$, para todo u tal que $(t,u) \in \mathscr{B}_i$. Ora, sabemos que a relação $\mathscr{B}$ é transitiva. Daí, pela definição 2.15-13, $\eta, s \models^{\Psi}_{T} B_i p$. Logo, $\eta, s \models^{\Psi}_{T} B_i L_i p \rightarrow B_i p$. Como s e η foram escolhidos arbitrariamente, concluímos a primeira parte da prova mostrando $\models B_i L_i p \rightarrow B_i p$. □

Demonstração. <u>Caso 2</u>. $\models B_i p \rightarrow B_i L_i p$.

Suponha um estado qualquer s de uma estrutura qualquer η tal que $\eta, s \models^{\Psi}_{T} B_i p$. Suponha agora, para redução ao absurdo, $\eta, s \models^{\Psi}_{F} B_i L_i p$. Assim, pela definição 2.15-14, $\eta, t \models^{\Psi \cap A_i(s)}_{F} L_i p$, para algum t tal que $(s,t) \in \mathscr{B}_i$. Como a relação $\mathscr{B}$ é serial, sabemos que há ao menos um u tal que $(t,u) \in \mathscr{B}_i$. Como a relação $\mathscr{B}$ também transitiva, $(s,u) \in \mathscr{B}_i$. Daí, aplicando a definição 2.15-13 na hipótese $\eta, s \models^{\Psi}_{T} B_i p$, obtemos $\eta, u \models^{\Psi \cap A_i(s)}_{T} p$, para qualquer u tal que $(t,u) \in \mathscr{B}_i$. Assim, novamente pela definição 2.15-13, $\eta, t \models^{\Psi \cap A_i(s)}_{T} B_i p$. Sabemos, através do item 4 do teorema 2.22, que $\eta, t \models^{\Psi \cap A_i(s)}_{T} B_i p \rightarrow L_i p$. Tendo isto, obtemos $\eta, t \models^{\Psi \cap A_i(s)}_{T} L_i p$. Dai, por 2.15-13, $\eta, s \models^{\psi}_{T} B_i L_i p$. Mas isso gera uma contradição com a hipótese anterior, a saber, $\eta, s \models^{\Psi}_{F} B_i L_i p$. Nossa hipótese para redução ao absurdo deve, então, ser falsa. Logo, $\eta, s \models^{\Psi}_{T} B_i p \rightarrow B_i L_i p$. Como s e η foram escolhidos arbitrariamente, concluímos com $\models B_i p \rightarrow B_i L_I p$. □

Se foi provado, ambos, $\models B_i L_i p \rightarrow B_i p$ e $\models B_i p \rightarrow B_i L_I p$, podemos introduzir $\leftrightarrow$ e concluir com o resultado desejado: $\models B_i L_i p \leftrightarrow B_i p$ □

Assim, na lógica da consciência, o agente acredita explicitamente que acredita implicitamente em uma fórmula p, exatamente se acredita explicitamente em p (FAGIN & HALPERN, 1988, p. 51).

Foi possível observar até então uma série de propriedades da lógica da consciência que concordam com a lógica das crenças explícitas e implícitas de Levesque, bem como outras em que tal fato não ocorre. Assim como em Levesque, a lógica da consciência permite, para as crenças implícitas, as três propriedades de onisciência lógica. Ou seja, as crenças implícitas são fechadas sob implicação, implicação válida, e os agentes acreditam implicitamente em todas as fórmulas válidas. Já em Levesque, as crenças explícitas não possuem quaisquer dessas três propriedades; isto é, as crenças explícitas não são fechadas sob implicação, nem sob implicação válida, além de os agentes não acreditarem necessariamente em todas as fórmulas válidas.

Como vimos, não é exatamente isso que ocorre na lógica da consciência, já que as crenças explícitas são fechadas sob implicação. Alguém poderia argumentar que isso é uma desvantagem que a lógica da consciência possui com relação à lógica de Levesque. Isso se sustentarmos, por exemplo, que o esquema $B_i(p \rightarrow q) \rightarrow (B_i p \rightarrow B_i q)$ nunca deva ser satisfeito, sob qualquer hipótese[58]. Apesar disso, a lógica da consciência possui uma série de propriedades interessantes. Um exemplo disso é o fato de $B_i(p \wedge \neg p)$ ser insatisfatível. Ou seja, os agentes não acreditam na incoerência. Essa é, inclusive, uma das críticas contra a lógica de Levesque, que sustenta por exemplo $(Bp \wedge B\neg p) \leftrightarrow B(p \wedge \neg p)$; isto é, o agente pode ter crenças inconsistentes se somente se toda situação que o agente considera possível for incoerente. Ora, tal propriedade não é amplamente aceita. Nem todos consideram possíveis situações incoerentes. Eis então algo que a lógica da consciência possui em vantagem com relação à lógica de Levesque; a lógica da consciência não precisa admitir a existência de situações incoerentes para invalidar o fecho sob implicação válida[59]. Outro detalhe importante é que – creio eu – epistemólogos que sustentam a validade do fecho sob implicação não achariam grandes incompatibilidades entre a lógica da consciência e sua tese de defesa do referido tipo de fecho. Como podemos constatar, a tese de Hawthorne da validade do fecho epistêmico (E-CLOS 1) é, por exemplo, plenamente compatível com aquilo que é modelado pela lógica da consciência.

Há, ainda, algumas diferenças entre a lógica de Levesque e a lógica da consciência que gostaríamos de frisar. É conveniente fazer menção, por exemplo, ao tratamento de fórmulas do tipo $B_i p$ e $L_i p$. Em Levesque, $\eta, s \vDash_F Bp$ se, e somente se, $\eta, s \nvDash_T Bp$ (o mesmo ocorre com $L_i p$). Ou seja, uma situação qualquer suporta a falsidade de uma crença explícita (ou implícita) exatamente se não suporta sua verdade. Na lógica da consciência, a condição de suporte para $B_i p$ e $L_i p$ funciona de modo diferente, a saber, $\eta, s \vDash_F^{\Psi} B_i p$ se, e somente se, $\eta, t \vDash_F^{\Psi \cap A_i(s)} p$, para algum t tal que $(s,t) \in \mathscr{B}_i$. Deste modo, uma estado s suporta a falsidade de uma fórmula qualquer $B_i p$ se e somente se há um outro estado t que o agente i considera possível tal que t suporta a falsidade de p. Como apontam Fagin & Halpern (1988, p. 51),

[58]Inclusive, já argumentei contra isso (seção 2.2.5), mostrando que em certos contextos do senso-comum é perfeitamente natural adotar o esquema $K_a(p \rightarrow q) \rightarrow (K_a p \rightarrow K_a q)$ como axioma.

[59]Ou seja, os "impossíveis mundos possíveis" de Hintikka não são levados em consideração na lógica da consciência.

"[...] isso significa que o agente tem de ter evidência positiva suportando a falsidade de $B_i p$, ao invés de apenas não ter evidência suportando a verdade de $B_i p$."[60]; isto é, a lógica da consciência não recorre à "falácia da ignorância" na relação de suporte em questão. Logo, pode-se afirmar que a lógica da consciência possui, nesse ponto, relações de suporte mais exigentes que as da lógica de Levesque.

No entanto, não há a necessidade de ficar comparando as duas lógicas, ora favorecendo uma, ora favorecendo outra. É plausível afirmar que as duas lógicas oferecem modelos razoáveis para a interpretação lógica da crença (e do conhecimento). Todas possuem pontos fortes e fracos. Porém, se é desejo do pesquisador da área da lógica da crença ou do conhecimento oferecer um modelo capaz de satisfazer todas as falhas de onisciência lógica, pode-se afirmar com certeza que nenhuma das duas é a mais adequada – principalmente porque, como sabemos, nunca podemos estar certos se todas as causas de falhas de oniscências nos são conhecidas, isto é, se todas já foram catalogadas por nós.

Há ainda várias outras tentativas de solução para o problema. Podemos então investigar qual delas é mais bem sucedida do que as outras (àquilo que se propõem), ou simplesmente atentarmos para o fato de que é possível construir lógicas particulares para a crença e para o conhecimento. Isto é, podemos nos fazer a seguinte pergunta: faz-se necessário construir lógicas gerais para a crença e conhecimento, de modo que sejam modelos capazes de representar todas as falhas de onisciência lógica? Ou será possível investigar a crença e o conhecimento através de vários modelos menores, cada um dando conta de aspectos particulares? Ou será possível ambos? Será que uma abordagem se sobrepõe a outra? Essa questão ainda será discutida neste trabalho. Responder essa questão é responder, por exemplo, se a abordagem de Hintikka é ou não de fato insuficiente.

2.3.3 Lógica da consciência geral

A lógica da consciência geral suporta um modelo que permite captar a falha de onisciência lógica também por recursos computacionais e de tempo limitados. Isso é possível dependendo da interpretação que se dá à noção de "consciência". Uma propriedade interessante da lógica da consciência geral é a de que, diferentemente da lógica da consciência, as crenças explícitas não são fechadas

[60]"[...] *this means that the agent has to have positive evidence supporting the falsity of* $B_i p$, *rather than just no evidence to support the truth of* $B_i p$."

sob implicação. Dizemos ser interessante porque a referida propriedade é possível sem a admissão dos estados incoerentes de Levesque, ou dos impossíveis mundos possíveis de Hintikka. Na lógica de Levesque, apesar de podermos introduzir um operador de consciência para a eliminação do fecho sob implicação, são as situações incoerentes que desempenham o papel central. Já a lógica da consciência consegue, sem admitir estados incoerentes, invalidar o fecho sob implicação válida e as crenças em fórmulas válidas, sucumbindo porém ao fecho sob implicação. Na lógica da consciência geral, veremos porque isso não ocorre.

Todavia, veremos também que ela é severamente criticada por Konolige (1986a), em várias passagens. A crítica de Konolige será discutida ao final desta seção. Iniciemos com a apresentação da lógica da consciência geral. Aqui, encontramos três operadores:

B: crenças explícitas

L: crenças implícitas

A: operador de consciência

Como Fagin & e Halpern apontam (1988, p. 52), podemos dar uma série de interpretações à fórmula $A_i p$: "i está ciente de p", "i está apto a descobrir a verdade de p" ou, quando se referindo a bases de conhecimento, "i está apto a computar a verdade de p no tempo T".

Definição 2.17. *Estrutura: uma estrutura de Kripke para a lógica da consciência geral é uma quádrupla* $\eta = (S, \pi, \mathscr{A}_i, ..., \mathscr{A}_n, \mathscr{B}_i, ..., \mathscr{B}_n)$. *Os membros de* η*:*

S: um conjunto de estados;

$\pi(s)$*: uma atribuição de verdade para cada estado* $s \in S$*;*

$\mathscr{B}_i$*: uma relação serial, transitiva e euclideana sobre S, para cada agente i.*

$\mathscr{A}_i(s)$*: um conjunto arbitrário de fórmulas (não somente de fórmulas primitivas, como é o caso na lógica da consciência).*

Neste momento, devemos nos deter um pouco para investigar certas características do operador A. Este operador é essencialmente sintático. Um agente i está ciente de φ se, e somente se, $\varphi \in \mathscr{A}_i(s)$, sendo $\mathscr{A}_i(s)$ o conjunto das fórmulas que o agente i está ciente no estado s. A ideia é quase a mesma da lógica da consciência. Porém, temos uma diferença: a noção de consciência não diz respeito apenas a fórmulas primitivas. Deste modo, um agente pode não acreditar explicitamente em φ por não está ciente de φ, sendo φ possivelmente uma fórmula composta.

Outra característica do operador A: flexibilidade. A lógica da consciência geral nos deixa muitas opções quando se tratando da interpretação que temos da ideia de consciência. É possível obtermos várias interpretações; isso vai depender da quantidade de restrições impostas à função de consciência. Fagin & Halpern apontam (1988, p. 53):

> A interpretação precisa que damos à noção de consciência dependerá da aplicação pretendida da lógica em questão. Colocando várias restrições sobre a função de consciência, podemos capturar várias noções distintas interessantes[61].

Se não há restrição à noção de consciência, é possível que, ambas, $\varphi \in \mathscr{A}_i(s)$ e $\neg\varphi \in \mathscr{A}_i(s)$. Assim, é possível que o conjunto $\mathscr{A}_i(s)$ seja inconsistente. É possível também que $\varphi \wedge \psi \in \mathscr{A}_i(s)$, mas $\psi \wedge \varphi \notin \mathscr{A}_i(s)$. Isso é possível pelo seguinte fato: as fórmulas das quais um agente está ciente não são necessariamente as que ele acredita. As diferentes restrições impostas sobre a função de consciência serão discutidas mais adiante.

Definição 2.18. *As condições de verdade da lógica da consciência geral são as seguintes:*

1. $\eta, s \models true$;
2. $\eta, s \models p$, *sendo* p *uma proposição primitiva, se e somente se* $\pi(s,p) = \textbf{\textit{true}}$;
3. $\eta, s \models \neg\varphi$ *se, e somente se,* $\eta, s \not\models \varphi$;
4. $\eta, s \models \varphi \wedge \psi$ *se, e somente se,* $\eta, s \models \varphi$ *e* $\eta, s \models \psi$;
5. $\eta, s \models A_i\varphi$ *se, e somente se,* $\varphi \in \mathscr{A}_i(s)$;
6. $\eta, s \models L_i\varphi$ *se, e somente se,* $\eta, t \models \varphi$, *para todo* t *tal que* $(s,t) \in \mathscr{B}_i$;

[61]"*The precise interpretation we give to the notion of awareness will depend on the intended application of the logic. By placing various restrictions on the awareness function, we can capture a number of interesting diferent notions.*" Novamente, há aqui uma posição favorável à tese da utilização de lógicas epistêmicas segundo interesses específicos de aplicação. Esta mesma postura, defendo eu, deve ser adotada no que se refere aos problemas epistemológicos gerados por alguns princípios de fecho epistêmico.

7. *$\eta, s \vDash B_i\varphi$ se, e somente se, $\varphi \in A_i(s)$ e $\eta, t \vDash \varphi$ para todo t tal que $(s,t) \in \mathscr{B}_i$.*

Como é possível observar, não há relações de suporte; as condições de verdade são obtidas apenas com a relação bivalorada padrão $\vDash$.

Alguns resultados na lógica da consciência geral

Há também outra particularidade que é importante frisar. O que define as crenças explícitas é a intersecção entre crenças implícitas e a função de consciência. Deste modo, um agente acredita explicitamente em uma fórmula φ se, e somente se:

1. O agente acredita implicitamente em φ;
2. O agente está ciente de φ.

Teorema 2.32. $\vDash B_ip \leftrightarrow A_ip \wedge L_ip$

Demonstração. Precisamos provar $\vDash B_ip \rightarrow A_ip \wedge L_ip$ e $\vDash A_ip \wedge L_ip \rightarrow B_ip$.

Demonstração. <u>Caso 1</u>: $\vDash B_ip \rightarrow A_ip \wedge L_ip$. Suponha $\vDash B_ip$. Assim, pela definição 2.18-7, $p \in \mathscr{A}_i(s)$ e $\eta, t \vDash p$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Ora, se $p \in \mathscr{A}_i(s)$, pela definição 2.18-5 obtemos $\eta, s \vDash A_ip$. Agora, por 2.18-6, se $\eta, t \vDash p$, para todo t tal que $(s,t) \in \mathscr{B}_i$, então $\eta, s \vDash L_ip$. Obtemos até então $\eta, s \vDash A_ip$ e $\eta, s \vDash L_ip$. Daí, por 2.18-4, segue-se $\eta, s \vDash L_ip \wedge A_ip$. Logo, $\vDash B_ip \rightarrow A_ip \wedge L_ip$. □

Demonstração. <u>Caso 2</u>. Desta vez, tomamos como hipóteses $\eta, s \vDash A_ip$ e $\eta, s \vDash L_ip$. Com base nas mesmas cláusulas utilizadas no caso 1, chega-se a $\vDash B_ip$ e portanto $\eta, s \vDash L_ip \wedge A_ip \rightarrow B_ip$. □

Daí, se foi provado $\vDash B_ip \rightarrow A_ip \wedge L_ip$ e $\vDash L_ip \wedge A_ip \rightarrow B_ip$, concluímos com $\vDash B_ip \leftrightarrow A_ip \wedge L_ip$. □

Este resultado mostra que é impossível para um agente da lógica da consciência geral ter crenças explícitas acerca de algo que ele não está ciente. Por conta disso, as crenças implícitas e explícitas irão confundir-se uma com a outra, se for assumido que o agente está ciente de todas as fórmulas.

Pela observação da cláusula 6 da definição 2.18, é fácil perceber que o operador L_i possui exatamente as mesmas propriedades do operador modal clássico para a crença[62]. Deste modo, o sistema modal clássico para a crença $KD45$ é um subsistema da lógica da consciência geral.

Teorema 2.33. *Na lógica da consciência geral, são válidas as seguintes propriedades:*

1. *A1: Todas as instâncias das tautologias proposicionais;*
2. *A2:* $(L_i p \wedge L_i(p \rightarrow q)) \rightarrow L_i q$;
3. *A3:* $\neg L_i(false)$;
4. *A4:* $L_i p \rightarrow L_i L_i p$;
5. *A5:* $\neg L_i p \rightarrow L_i \neg L_i p$;
6. *R1: Se* p, $p \rightarrow q$, *então* q;
7. *R2: Se* p, *então* $L_i p$.

Demonstração. As provas são similares às que foram feitas na lógica da consciência, bastando apenas fazer as devidas adaptações. □

Se for de nosso interesse transformar a lógica da consciência geral em um sistema de conhecimento, ao invés de crença, assumimos a relação B_i como sendo de equivalência, e adicionamos o esquema $L_i p \rightarrow p$ à nossa lista de axiomas.

Também na lógica da consciência geral, ainda assim a fórmula $\neg B_i(p \vee \neg p)$ é satisfatível. Ou seja, nem sempre os agentes acreditam explicitamente em todas as fórmulas válidas. Isso ocorre por ser possível que o agente não esteja ciente de $p \vee \neg p$. Como foi visto que o ato de estar ciente é determinante para se ter uma crença explícita, e o agente não está ciente de $p \vee \neg p$, concluímos que ele não pode acreditar explicitamente em $p \vee \neg p$; daí, $\neg B_i(p \vee \neg p)$.

Vale relembrar o seguinte: a função de consciência não se aplica apenas às fórmulas primitivas, de modo que um agente pode não acreditar explicitamente em $p \vee \neg p$ simplesmente por não estar ciente de $p \vee \neg p$; isto é, para estar ciente ou não de uma fórmula, não é necessário que o agente esteja ciente ou não acerca de

[62] Também é fácil perceber também que $B_i p \rightarrow L_i p$.

suas partes. Melhor dizendo, a noção de consciência não precisa necessariamente ser fechada sob sub-fórmulas, dado que $\mathscr{A}_i(s)$ é um conjunto arbitrário. Todavia, se nos interessa impor o fecho por sub-fórmulas à noção de consciência, não é difícil fazê-lo.

Continuemos com mais alguns resultados. Até agora, a satisfatibilidade da fórmula $\neg B_i(p \vee \neg p)$ e o comportamento clássico do operador L_i são, ambos, também resultados da lógica da consciência. Porém, eis um resultado que diferencia uma lógica da outra: o fecho sob implicação das crenças explícitas, isto é, o princípio (E-CLOS 1). Do modo como a lógica da consciência é construída, a saber, sem os estados incoerentes, ela fica sujeita ao fecho sob implicação das crenças explícitas. Na lógica da consciência geral, mesmo sem os estados incoerentes, o princípio (E-CLOS 1) para as crenças explícitas não é uma propriedade válida.

É possível para um agente acreditar explicitamente em p e $p \to q$ sem, necessariamente, acreditar explicitamente em q. Ou seja, a fórmula $(B_i p \wedge B_i(p \to q)) \wedge \neg B_i q$ é satisfatível: o agente pode não estar ciente de q; se não está ciente de q, então não pode acreditar explicitamente em q. Dado que $\mathscr{A}_i(s)$ é um conjunto arbitrário, é possível que ambos $p \in A_i(s)$ e $p \to q \in \mathscr{A}_i(s)$, mas $q \notin \mathscr{A}_i(s)$. A função de consciência não precisa ser fechada sob sub-fórmulas! Fagin & Halpen (1988, p. 54), contudo, apresentam uma série de restrições opcionais à noção de consciência, incluindo o fecho sob sub-fórmulas:

1. É possível impor a seguinte restrição à noção de consciência: $\varphi \wedge \psi \in \mathscr{A}_i(s)$ se, e somente se, $\psi \wedge \varphi \in \mathscr{A}_i(s)$. Isto é, a ordem dos conjuntos é relevante com relação à noção de consciência. Se isso for o caso, então obteremos $\vDash (p \wedge q) \leftrightarrow (q \wedge p)$. Esta restrição pode ser aplicada, de maneira similar, à negação. Isto é, a noção de consciência poderia ser de tal modo que um agente está ciente de uma fórmula φ se, e somente se, está ciente de sua negação; assim, $\varphi \in \mathscr{A}_i(s)$ se, e somente se, $\neg\varphi \in \mathscr{A}_i(s)$.

2. Fecho sob sub-fórmulas: Se $\varphi \in \mathscr{A}_i(s)$ e ψ é uma sub-fórmula de φ, então $\psi \in \mathscr{A}_i(s)$;

3. Aplicação da noção de consciência apenas com relação às fórmulas primitivas. Deste modo, o agente pode estar ciente apenas de certas fórmulas primitivas de um conjunto Ψ qualquer. O conjunto $\mathscr{A}_i(s)$ consiste, então, de proposições primitivas do conjunto Ψ. Esta é, por exemplo, a noção encontrada na lógica

da consciência. Se a referida condição for imposta, a lógica da consciência e a lógica da consciência geral terão muitas coisas em comum, com algumas exceções. Na primeira, $\models B_i p \rightarrow B_i(p \vee q)$, enquanto na segunda $\not\models B_i p \rightarrow B_i(p \vee q)$.

4. Além de trabalhar com a consciência de fórmulas primitivas, é possível também trabalharmos com a consciência com relação a agentes. Deste modo, um agente qualquer j pode ou não estar ciente acerca de alguma fórmula mencionada por um agente i.

5. Um agente pode ser auto-reflexivo. Isto é, se $\varphi \in \mathscr{A}_i(s)$, então $\models A_i\varphi \in \mathscr{A}_i(s)$. Se essa condição for imposta, então $A_i p \rightarrow A_i A_i p$. Daí, como resultado, $\models B_i p \rightarrow B_i B_i p$. Ou seja, se a condição 5 for imposta, a lógica da consciência passa a satisfazer a regra $(C.KK^*)$ de Hintikka, para as crenças explícitas[63].

6. Esta restrição é parecida com a anterior; é possível para um agente saber de quais fórmulas ele está ciente. Isto é, se $(s,t) \in \mathscr{B}_i$, então $\mathscr{A}_i(s) = \mathscr{A}_i(t)$. Se esta condição for imposta, então $\models A_i p \rightarrow L_i A_i p$ e $\models \neg A_i p \rightarrow L_i \neg A_i p$.

7. Nesta restrição, os elementos do conjunto $\mathscr{A}_i(s)$ consistem daquelas fórmulas que o agente i consegue determinar, em algum tempo ou espaço limitado, se seguem ou não da informação de que dispõe no estado atual s.

Com a flexibilidade da função de consciência, a lista e os tipos de teoremas irão depender da quantidade de restrições que houvermos imposto. Cada escolha de restrições resulta em uma axiomatização diferente, e portanto em lógicas diferentes. Em certo momento podemos aceitar as condições 5 e 7, e rejeitar as demais; em outro momento podemos aceitar 1 conjuntamente com 7, e rejeitar 5, etc. Esse exemplo serve para mostrar que o agente que trabalha com a própria lógica da consciência (para modelar crenças e conhecimentos de outros agentes) pode variar no modo como usa o raciocínio, de acordo com a situação. Não precisamos, portanto, nos agarrarmos a qualquer noção absoluta sobre o ato de estar ciente. É interessante conhecermos quais são todos os recursos possíveis de serem utilizados, sem

[63]Ao começar o trabalho com a abordagem de Hintikka, foi pensado também em se acrescentar uma discussão sobre a regra $(C.KK^*)$. Na verdade, o princípio da introspecção positiva é mais investigado em *Knowledge and Belief* do que o próprio problema da onisciência lógica. Basicamente, todo o capítulo 4 da referida obra é dedicado à discussão da regra $(C.KK^*)$, no qual Hintikka irá apresentar vários argumentos presentes na História da Filosofia que advogam em favor de sua regra.

necessariamente utilizarmos todos ao mesmo tempo. Novamente, aqui, as circunstâncias dirão que lógica melhor expressa os agentes envolvidos no contexto.

Agora, vejamos outras propriedades interessantes. Já foi dito que é possível demonstrar todos os esquemas clássicos para o operador L_i; isto é, KD45 é um subsistema da lógica da consciência geral (com relação apenas ao operador L_i). É interessante mostrar também como a lógica da consciência geral é axiomatizável; Fagin & Halpern apontam (1988 p. 54) que, para isso, basta apenas adicionar $B_i p \leftrightarrow L_i p \wedge A_i p$ aos axiomas de KD45[64]. Chamemos, por praticidade, esse esquema de A6. Ao adotá-lo, é possível demonstrarmos sintaticamente:

1 . $(B_i p \wedge B_i(p \rightarrow q) \wedge A_i q) \rightarrow B_i q$; ou seja, se um agente acredita explicitamente em p e $p \rightarrow q$, então o agente acreditará em q tão logo esteja ciente de q.

2 . $A_i p \rightarrow B_i p$; o agente deve estar ciente da fórmula relevante antes de acreditar explicitamente na mesma.

Se a restrição 6 for adotada para a função de consciência, então:

3 . $B_i p \wedge A_i B_i p \rightarrow B_i B_i p$; representação para a lógica da consciência geral da introspecção positiva aplicada ao operador de crença explícita.

4 . $\neg B_i p \wedge A_i \neg B_i p \rightarrow B_i \neg B_i p$; representação para a lógica da consciência geral da introspecção negativa aplicada ao operador de crença explícita.

Os itens 3 e 4 são particularmente interessantes; por conta disso, Fagin & Halpern (1988 p. 55) dão atenção especial a esses dois esquemas:

> Novamente, note que um agente deve estar ciente acerca da fórmula relevante antes que ele acredite nela explicitamente. O primeiro desses dois axiomas [itens 3 e 4] mostra, como na citação de Chardin, um animal pode saber algo, mas sem saber que sabe, enquanto que a segunda indica como um agente pode ser tão "estúpido que ele nem mesmo sabe que ele não sabe de p"[65].

[64] Apesar de apontarem também que isso não nos adiciona qualquer novo *insight* sobre as propriedades das crenças explícitas.

[65] *"Again, note that an agent must be aware of the relevant formula before he explicitly believes it. The first of these two axioms shows how, as in the quote from de Chardin, an animal may know, but not know that it knows, while the second indicates how and agent may be "so dumb that he doesn't even know that he doesn't know p."*"

Além dos esquemas de introspecção, há também outros menos questionáveis (inclusive muito comuns na lógica proposicional clássica). Um exemplo é o esquema acerca da ordem dos conjuntos. Deste modo, se a restrição 1 for adotada sobre a função de consciência, teremos:

5 . $B_i(p \wedge q) \leftrightarrow B_i(q \wedge p)$; sintaticamente, este esquema é derivado utilizando o axioma $A_i(p \wedge q) \leftrightarrow A_i(q \wedge p)$ (chamemo-lo de A7). Vale ressaltar que o axioma A7 pode ser utilizado se, e somente se, a restrição número 1 sobre a noção de consciência for adotada.

Da mesma forma, ainda com a restrição 1, é possível derivar:

6 . $B_i p \leftrightarrow B_i \neg\neg p$; para que isso ocorra, basta utilizar o axioma A8 $A_i p \leftrightarrow A_i \neg p$, que segue da restrição 1.

Crítica de Konolige à lógica da consciência geral

Antes de tudo, vale ressaltar que a lógica da consciência geral, e sua estratégia de utilizar uma função psicológica como "consciência" de um ponto de vista sintático é bastante perspicaz. Ela combina a abordagem padrão de mundos possíveis com elementos da abordagem sintática. O que a destaca da lógica da consciência é a completa arbitrariedade do conjunto de consciência, que não se limita apenas às proposições primitivas. Essa estratégia – isto é, de combinar a semântica de mundos possíveis e a noção sintática da função de consciência – como foi colocado por Fagin & Halpen em sua versão preliminar da lógica da consciência geral, "preserva a elegância e o apelo intuitivo da abordagem semântica"[66]. Contudo, Konolige, que escreveu sua tese de doutorado (KONOLIGE, 1984) propondo uma abordagem essencialmente sentencial para lidar com o problema, criticou severamente a lógica da consciência geral (KONOLIGE, 1986a), argumentando que ela é apenas uma versão mais complicada da abordagem sentencial.

Por exemplo, foi visto que o conjunto de fórmulas das quais um agente está ciente é apenas uma lista de fórmulas. Em primeiro momento, é apenas uma sequência de símbolos, de modo que é possível para um agente estar ciente de uma fórmula φ sem estar ciente de uma fórmula ψ, mesmo que φ e ψ sejam

[66] R. FAGIN & J. Y. HALPERN. **Belief, Awareness and Limited Reasoning.**. *In*: Ninth International Joint Conference on AI. Los Angeles, CA: 1985. p. 491-501.

logicamente equivalentes. Em vista disso, encontramos a seguinte passagem em Konolige (1986a, p. 242):

> Aqui nós tomamos uma visão crítica dessa posição, e argumentamos que a lógica da consciência geral é essencialmente equivalente à abordagem sentencial, e a adição de mundos possíveis gera complicações sem motivos e contra-intuitivas[67].

Para entender plenamente o argumento que Konolige oferece, seria preciso apresentarmos um sistema sintático de crenças (de preferência o seu próprio, em sua tese de doutorado). Sendo assim, o próximo capítulo lida com a abordagem sentencial oferecida por Konolige. Contudo, achamos conveniente explorar primeiro as abordagens que ainda utilizaram, mesmo que de maneira questionável (como é o caso da lógica da consciência geral), o modelo semântico de mundos possíveis[68].

Enquanto a abordagem sintática de Konolige não é vista de modo completo, faz-se necessário, pelo menos, uma caracterização simples de seu sistema. O sistema lógico-dedutivo que Konolige apresenta para as crenças consiste, grosso modo, de um conjunto básico de crenças, e um certo tipo de mecanismo formal (regras dedutivas) para derivação de novas crenças a partir deste conjunto. O conjunto das crenças implícitas é o conjunto das consequências lógicas das crenças explícitas. Deste modo, em uma lógica sentencial, o objetivo é entender como os agentes derivam sintaticamente uma crença a partir de outra.

Retomando a crítica contra a lógica da consciência geral, observamos que Konolige (1986a p. 245) mostra que as crenças explícitas são definidas a partir da intersecção das crenças implícitas e a função de consciência. Não é difícil mostrar, então:

Teorema 2.34. *Se um agente está ciente de todas as fórmulas, então* $\models (B_i p \wedge B_i(p \rightarrow q)) \rightarrow B_i q$.

Demonstração. Suponha uma estrutura η qualquer e um estado s qualquer desta mesma estrutura. Suponha que $\eta, s \models B_i p$ e $\eta, s \models B_i(p \rightarrow q)$. Assim, pela definição 2.18-7, $p \in \mathscr{A}_i(s)$, $p \rightarrow q \in \mathscr{A}_i(s)$, $\eta, t \models p$ e $\eta, t \models p \rightarrow q$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Como $\models$ é uma relação bivalorada padrão, é possível definir a implicação do

[67] *"Here we take a critical view of this position, and argue that a logic of general awareness is essentially equivalent to the sentencial approach, but the additon of possible-worlds elements adds unmotivated and unintuitive complications."*

[68] De fato, apresentando nessa ordem, é possível observar com clareza a proximidade de todas as abordagens com o modelo pioneiro de Hintikka.

mesmo modo como em lógica clássica. Daí, obtemos sem problemas $\eta,t \vDash q$ a partir de $\eta,t \vDash p$ e $\eta,t \vDash p \to q$ (sendo q o caso em todo t, $(s,t) \in \mathscr{B}_i$). Foi tomado como hipótese que o agente está ciente de todas as fórmulas relevantes. Sendo assim, $q \in \mathscr{A}_i(s)$. Então, $q \in \mathscr{A}_i(s)$ e $\eta,t \vDash q$, para todo t tal que $(s,t) \in \mathscr{B}_i$. Logo, por 2.18-7, $\eta,s \vDash B_i q$. Como a estrutura η e o estado s são ambos arbitrários, concluímos com $\vDash (B_i p \wedge B_i(p \to q)) \to B_i q$. $\square$

Por conta disso, Konolige diz (1986a p. 245):

> A lógica da consciência geral caracteriza portanto os agentes como raciocinadores perfeitos, restritos a considerarem um subconjunto de sentenças possíveis para raciocinar sobre elas[69].

Ou seja, os agentes da lógica da consciência são raciocinadores perfeitos, isto é, com capacidades computacionais infinitas; porém, com possíveis limitações quantitativas sobre as fórmulas das quais eles estão cientes. A partir daí, duas questões importantes são levantadas por Konolige (1986a, p. 246):

1. A semântica de mundos possíveis desempenha um papel fundamental nessa abordagem?

2. O quão intuitiva e útil é a noção geral de raciocinadores perfeitos (com limitações acerca da consciência de certas fórmulas) para representar o raciocínio limitado?

A resposta à primeira pergunta é negativa. Konolige mostra que, quando se trata de consciência, a conexão entre as condições de acessibilidade e os conjuntos de consciência (isto é, os conjuntos das fórmulas das quais um agente está ciente em um dado estado) é quebrada; por conta disso, a própria conexão que há entre as condições de acessibilidade e as crenças dos agentes também é quebrada. Observemos a seguinte demonstração:

[69] *"The logic of general awareness thus caracterizes agents as perfect reasoners, restricted in some way to considering a subset of possible sentences to reason about."*

1. $B_i p$	Hipótese
2. $B_i p \leftrightarrow L_i p \wedge A_i p$	A6
3. $L_i p$	1,2 MP
4. $A_i p$	1,2 MP
5. $L_i L_i p$	3, Necessitação
6. $A_i p \rightarrow L_i A_i p$	Restrição 6
7. $L_i A_i p$	4,6 MP
8. $B_i p \rightarrow A_i B_i p$	Cons. Restrição 5
9. $A_i B_i p$	1,8 MP
10. $L_i L_i p \wedge L_i A_i p$	5,7 $\wedge I$
11. $L_i L_i p \wedge L_i A_i p \wedge A_i B_i p$	9,10 $\wedge I$
12. $(L_i B_i p \wedge A_i B_i p) \leftrightarrow (L_i L_i p \wedge L_i A_i p \wedge A_i B_i p)$	Teorema
13. $L_i B_i p \wedge A_i B_i p$	11,12 MP
14. $B_i B_i p \leftrightarrow (L_i B_i p \wedge A_i B_i p)$	Teorema
15. $B_i B_i p$	13,14 MP

Está claro que a fórmula $B_i B_i p$ é derivada a partir de $B_i p$. É possível fazer tal derivação sem que os esquemas essenciais à demonstração (mais precisamente, os da linha 6 e 8) sejam afetados pelas estruturas de acessibilidade. Sendo assim, no presente caso – que inclusive representa uma derivação importante para as crenças explícitas – as estruturas de acessibilidade não têm valor algum. Como o próprio Konolige coloca (KONOLIGE, 1986a, p. 247):

> Logo, a bela análise formal das propriedades introspectivas obtíveis na semântica de Kripke não está presente na lógica da consciência geral[70].

Uma outra objeção é feita. De acordo com Konolige, também é possível mostrar que a lógica da consciência geral possui uma semântica sentencial natural. Para isso, ele faz uma observação sobre a cláusula 7 da definição 2.18:

- $\eta, s \models B_i \varphi$ se $\varphi \in A_i(s)$ e $\eta, t \models \varphi$ para todo t tal que $(s,t) \in \mathscr{B}_i$.

Esta é a definição de crença explícita, que consiste na intersecção entre crença implícita e consciência. Claramente, a definição consiste de uma conjunção. Konolige aponta que o primeiro lado desta conjunção "[...] refere-se explicitamente a um conjunto de sentenças" (KONOLIGE, 1986a, p. 247)[71]; já a segunda metade

[70]"*Hence the nice formal analysis of introspective properties obtainable in Kripke semantics is not present in the logic of general awareness.*"

[71]"[...] *refers explicitly to a set of sentences.*"

faz uso dos mundos possíveis. A intenção de Konolige consiste, portanto, em mostrar que também o segundo lado da definição pode ser posto em uma caracterização sentencial. Para que isso seja possível, basta lembrar que o operador L da lógica da consciência geral é axiomatizável como **KD45** + consistência. Daí, Konolige se utiliza do resultado de Moore[72], que mostrou que o fraco **KD45** caracteriza o que ele chama de *conjuntos estáveis*. Seguindo ainda Konolige, um conjunto S é considerado estável se ele é:

1. Fechado sob *modus ponens*;
2. Contém todas as tautologias;
3. Obedece as condições: 1. Se $\varphi \in S$, $L\varphi \in S$; 2. Se $\varphi \notin S$, $\neg L\varphi \in S$.

Não é difícil perceber que o sistema proposto na lógica da consciência geral possui todas as propriedades enumeradas acima. Sendo assim, "de um ponto de vista sentencial, os modelos da lógica da consciência geral consistem de conjuntos estáveis em intersecção com um conjunto arbitrário de consciência." (KONOLIGE, 1986a, p. 247).

O que dizer então da segunda questão? Uma pergunta pertinente levantada a partir deste fato é acerca da motivação que está por trás dessa lógica. Afinal, o modelo proposto pela lógica da consciência geral está em consonância com a noção psicológica comum que se têm sobre as crenças? De certo modo, Konolige (1986a, p. 247) reconhece que há pontos positivos:

> [...] o modelo dedutivo parece plausível em sua forma geral, porque os agentes aprendem fatos sobre o mundo a partir de suas observações ou por terem sido informados por terceiros, e então continuam deduzindo novas consequências. Podemos esperar caracterizar as crenças de um agente com recursos limitados notando que fatos ele aprende, e que regras ele usa para inferir outros fatos. Tal modelo seria útil para predizer o comportamento de um agente, dado informações parciais sobre suas crenças[73].

Em contrapartida, os pontos negativos são claros :

[72] MOORE, Robert C. **Semantical considerations on nonmonotonic logic**. Technical note 284. Menlo Park: SRI International, 1983.

[73] "[...] *the deduction model seems plausible in its general form, because agents learn facts about the world form their observations or from been told, and then go on to deduce further consequences. We might expect to caracterize the beliefs of a resource-limited agent by noting what facts he learns, and what rules he uses to infer other facts. Such a model would be useful in predicting the behavior of the agent, given partial information about his beliefs.*"

> [...] a lógica da consciência geral representa os agentes como raciocinadores perfeitos, restritos a considerarem algumas classes sintáticas de sentenças [no caso, as classes de sentenças das quais o agente está ciente]. Não parece haver qualquer intuição clara de que isso é o caso para os agentes humanos ou artificiais[74].

Numa teoria sintático-semântica para a crença (ou conhecimento) na qual o ato de estar ciente desempenha um papel fundamental, é possível distinguir ao menos duas abordagens possíveis (KONOLIGE, 1986a p. 248):

1. Os agentes computam todas as consequências lógicas de suas crenças, jogando fora aquelas que não estão no conjunto de consciência (aqui Konolige sugere que isso ocorre talvez por falta de memória); todavia, ele não acha que esse modelo seja suficiente, porque não faz referência também à falha de onisciência lógica por recursos limitados de tempo[75].

2. Os agentes usam um sistema lógico-dedutivo completo para computar as consequências de suas crenças, mas não perseguem aquelas linhas de raciocínio que requerem derivar sentenças fora do conjunto de consciência. Isto é plausível, porque as derivações incompletas dessa natureza poderiam ser alcançadas com limitações nos recursos de espaço e de tempo, dadas restrições fortes sobre as sentenças deriváveis. Contudo, isso é justamente o modelo dedutivo da crença.

O argumento de Konolige é, de fato, interessante. Ele mostra que a lógica da consciência geral, apesar da tentativa de manter as características da semântica de mundos possíveis, não escapa a uma formulação sentencial; deste modo, nada tem a acrescentar em relação a outras lógicas sentenciais já existentes, dado que pode ser formulada nos mesmos termos. Apesar disso, é preciso tomar cuidado contra mal-entendimentos:

> A lição que podemos tirar disso não é a de que as atitudes proposicionais e a visão sentencial das crenças sejam irreconciliáveis e por

[74]"[...] *the logic of general awareness represents agents as perfect reasoners, restricted to considering some sintatic class of sentences. There don't seem to be any clear intuitions that this is the case for human or computer agents.*"

[75]Ainda neste capítulo, argumentarei, de um certo modo, contra essa ideia de Konolige. Será defendido que os recursos limitados de tempo não devem constituir os subsídios teóricos principais na construção de uma teoria lógica da crença ou do conhecimento. Essa ideia é baseada em um argumento que será proposto aqui com o intuito de trivializar a falha de onisciência lógica por recursos limitados de tempo. Este, por sua vez, será chamado de *argumento do assassino 47*. Tal argumento tem como finalidade mostrar que nunca se sabe de quanto tempo um agente dispõe, e portanto não faz sentido ser a falta de tempo algo fundamental para a falha da onisciência lógica.

> isso não deveriam ser mixadas, mas sim que a introdução de mundos possíveis deva ser repensada[76]. (KONOLIGE, 1986, p. 248)

Vale também ressaltar que o argumento de Konolige se aplica a uma noção particular de consciência, conhecida como *consciência por computação*. Para o ponto de vista computacional, há duas abordagens para a noção de consciência. A primeira delas (item 1 acima), interpreta a função de consciência como um *filtro*; a segunda (item 2), interpreta a função de consciência como um *derivador*. Mas existem abordagens diferentes, que identificam a *consciência por percepção*. Nessa abordagem, estar ciente ou consciente de algo significa perceber algo. O ato de estar ciente acerca de algo composto é geralmente construído a partir da consciência de suas partes; ou seja, perceber um composto é perceber suas partes. Segundo Huang & Kwast (1991, p. 9), uma semântica apropriada para formalizar a noção de consciência por percepção é proposta por Barwise & Perry[77].

2.4 Abordagens sentenciais e onisciência lógica

A teoria lógica a seguir é uma abordagem essencialmente sentencial para o problema da onisciência lógica (KONOLIGE, 1984, 1986b). A palavra "sentencial" não é utilizada por acaso:

> A inteligência artificial se utilizou de duas tradições filosóficas diferentes para sua formalização da crença e do conhecimento. Em uma tradição, a crença é uma relação entre um agente e uma proposição, isto é, crença é uma *atitude proposicional*. Na outra tradição, a crença é uma relação entre um agente e uma sentença que expressa uma proposição. Chamaremos essa de abordagem *sentencial*[78]. (KONOLIGE, 1986a, p. 241).

Abordagens como a da semântica de mundos possíveis tendem, em princípio, a ganhar mais atenção do que as abordagens sentenciais. Segundo Konolige (1986, p. 241), isso acontece porque o modelo dos mundos possíveis facilita a análise lógica. De qualquer modo, o modelo dedutivo de crenças que o próprio Konolige

[76]*"The lesson we should draw from this is not that propositional attitude and sentencial views are irreconciliable and should not be intermixed, but rather that the introduction of possible worlds bears rethinking."*

[77]BARWISE, John; PERRY, John. Situations and attitudes. **Journal of Philosophy**. Vol 78, n. 11, p. 668-691, nov. 1981.

[78]*"Artificial Intelligence has borrowed from two different philosophical traditions for its formalizations of belief and knowledge. In one tradition, belief is a relation between an agent and a proposition, that is, belief is a propositional atttitude. In the other tradition, belief is a relation between an agent and a sentence that expresses a proposition. We will call this the sentential approach."*

apresenta (KONOLIGE, 1984, 1986b) fornece um aparato interessante para lidar com o problema da onisciência lógica. Com relação à abordagem sentencial, Whitsey escreve:

> [...] ela é, ao meu ver, o mais simples e mais intuitivo ponto de partida pelo qual os modelos para o pensamento limitado podem ser construídos (WHITSEY, 2003, p. 22)[79].

De fato, a abordagem sentencial tem um apelo muito intuitivo. Apresentemos, então, o modelo dedutivo de crenças (KONOLIGE, 1984).

2.4.1 A ideia do sistema dedutivo de crenças

A discussão inicial de Konolige acerca do sistema dedutivo de crenças é, em princípio, toda voltada para a inteligência artificial. A meta inicial do sistema dedutivo de crenças (KONOLIGE, 1984) é permitir a construção de um modelo capaz de examinar como *AI robots planning systems* representam e pensam sobre o mundo. Isso é feito com base na ideia do subsistema de crenças. Konolige (1984, p. 19) identifica as características principais desse sistema básico de crenças:

1. Um subsistema de crenças contém uma lista de sentenças de alguma linguagem interna ("mental"). Essas são as crenças básicas. As crenças básicas são responsáveis por todas as outras crenças que venham a surgir;

2. Os agentes podem inferir consequências de suas crenças pela manipulação sintática de sentenças do subsistema de crenças; esse processo é feito com base em regras de dedução e uma estratégia de controle (da prova);

3. A derivação das consequências das crenças é incompleta, devido a certas limitações no processo de inferência. Essa limitação pode acontecer, no mínimo, devido a três motivos:

   (a) As regras de inferência que o agente manipula podem ser fracas demais, de um ponto de vista lógico;

   (b) O agente pode decidir que algumas crenças não são relevantes para uma dada questão;

---

[79] "[...] *it is, to my mind, the simplest and most intuitive starting point from which models of resource-bounded reasoning can be built.*"

(c) A estratégia de controle do agente pode efetuar apenas um subconjunto de todas as derivações possíveis quando confrontado com alguma limitação de recursos.

Toda a ideia do sistema dedutivo de crenças se baseia, grosso modo, nesses critérios. A partir deles, é possível construir uma estrutura matemática que descreva, relativamente a esse modelo, o modo como os agentes derivam crenças e aprendem fatos sobre o mundo externo. O objeto matemático construído para essa teoria, com base nos critérios apresentados acima, é chamado de *estrutura de dedução*. Os componentes da *estrutura de dedução* serão apresentados mais adiante. No momento, nos deparamos com a seguinte questão: como funciona esse sistema dedutivo?

O sistema dedutivo é caracterizado como *subsistema de crenças*, que por sua vez é parte do *sistema de planejamento robô*[80]. O termo *sistema de planejamento* é utilizado devido ao seguinte fato: o agente representa o mundo externo e então planeja ações para atuar nesse mundo. Konolige reconhece (KONOLIGE 1984, p. 18) que é impossível representar o mundo real, pelo fato de ser extremamente complexo. Deste modo, os agentes consideram apenas descrições parciais da realidade. O *sistema de planejamento* lida, portanto, com uma abstração. Um estado desse mundo abstrato é conhecido como uma *situação*. Segue-se então que o *sistema de planejamento* terá apenas um conhecimento parcial de uma dada situação.

2.4.2 O processo de obtenção de crenças

Considerando que as crenças dos agentes são sentenças de um dado conjunto (o conjunto de crenças), temos que as referidas crenças são obtidas através de uma operação puramente sintática sobre as sentenças desse conjunto. As sentenças obtidas a partir do conjunto de crenças são chamadas de *sentenças derivadas*. O *subsistema de crenças* é constituído de duas partes:

1. Representação (na qual o agente organiza toda sua informação disponível sobre o mundo);

2. Dedução dos fatos acerca desse mundo (como já foi dito, isso é feito por meio de um processo sintático). O processo de dedução é, por sua vez, composto de:

[80] *Robot planning system.*

(a) Regras de dedução;

(b) Estratégia de controle; essa estratégia irá determinar quais regras de dedução serão utilizadas, como serão utilizadas e para onde irá a resposta quando uma dada pergunta é feita ao subsistema de crenças.

O subsistema de crenças opera do seguinte modo: o processo de dedução é uma operação sintática, tomando uma dada sentença como *input*, e produzindo novas sentenças como *output* (KONOLIGE, 1984, p. 18). Existem dois tipos de estímulos que resultam em uma reação por parte do sistema. Um deles consiste em deletar ou adicionar sentenças ao conjunto básico de sentenças; esse processo é conhecido como *atualização* ou *revisão* de crenças[81]. O outro tipo de estímulo consiste em fazer uma dada pergunta ao sistema. A pergunta, por sua vez, consiste em saber se uma dada sentença pertence ou não ao conjunto de crenças do sistema; isto é, se a referida sentença é ou não uma crença. Isso leva o sistema a tentar provar a sentença com base nas regras de dedução, e guiado pela estratégia de controle. Caso o sistema encontre uma prova para a sentença, a resposta será positiva; do contrário, negativa. Konolige sustenta (1984, p. 19-20) que todo esse processo de pergunta–resposta deva ser finito, e de preferência que ocorra em um intervalo curto de tempo:

> Se imaginamos um subsistema de crenças como parte de um agente robô que deve interagir com um ambiente em constante mudança, então a quantidade de tempo que o agente pode gastar computando consequências de suas crenças é extremamente limitada: como os agentes humanos, os robôs terão muitas vezes que agir rapidamente para responder a uma dada situação, sem poder dar-se ao luxo de ter recursos ilimitados para derivar um plano. Deste modo, uma importante propriedade do processo de pergunta–resposta é que ele sempre termina em uma quantidade finita (e geralmente pequena) de tempo[82].

Vale frisar que há dois tipos de perguntas possíveis para se fazer ao sistema, em espera a uma resposta. Citando Chang & Lee (1973), Konolige faz a distinção entre as perguntas do tipo A e as do tipo B:

[81]Konolige (1984, p. 19) reconhece que esse é um tema muito complexo por si só, e por isso não o considera nesse modelo de crenças.

[82]*"If we envision a belief subsystem as part of a robot agent that must interact with changing enviroment, then the amount of time the agent can spend computing consequences of its beliefs is strictly limited: like human agents, robot agents will often have to act quickly to respond a situation, without the luxury of unlimited resources for deriving a plan. Thus an important property of the belief query process is that it must always terminate in a finite (and usually small) amount of time."* (KONOLIGE, 1984, p. 20)

- A: Questão que envolve uma simples resposta de "sim" ou "não". Exemplo: "Está chovendo lá fora?", "O céu é azul?" ou "Renato Russo foi um filósofo?" etc; para responder a pergunta "Renato Russo foi um filósofo?", por exemplo, o subsistema simplesmente tenta derivar a sentença "Renato Russo foi um filósofo." a partir das sentenças básicas. Se o sistema conseguir a prova, a resposta é "sim"; caso contrário, a resposta é "não".

- B: Questão que envolve como resposta um indivíduo, conjunto de indivíduos ou condições que satisfazem uma dada propriedade. Exemplo: "Quais são os terroristas mais procurados do mundo?" etc. Para perguntas como essas, existem vários tipos de respostas; às vezes, a resposta pode ser o nome de um indivíduo (ex: Osama Bin Laden); em outros casos, a resposta pode ser a descrição de um indivíduo (ex: o líder da Al-Quaeda) ou de indivíduos (ex: os membros da Al-Quaeda). Como aponta Konolige (1984, p. 20), a resposta pode ser também uma disjunção (ex: Osama ou Bush).

2.4.3 As estruturas de dedução

Antes de apresentar a definição formal de *estrutura de dedução*, faz-se necessário apresentarmos formalmente o que já foi discutido até então. O conjunto básico de crenças de um agente i qualquer é simbolizado como KB_i. KB vem de *knowledge base*. A ideia é que KB_i represente o conjunto básico de crenças nas quais o agente acredita e das quais irá derivar outras crenças. O conjunto de regras de dedução de um agente é representado pela função p. Para cada agente i, $p(i)$ atribui um conjunto de regras de dedução a i. Escreve-se $\Gamma \vdash_{p(i)} \phi$ para dizer que a sentença ϕ é derivável a partir do conjunto de sentenças Γ, usando as regras atribuídas por $p(i)$. Se o agente i acredita em ϕ, escreve-se $B_i\phi$. O agente i acredita em ϕ se, e somente se:

1. $\phi \in B_i$; isto é, ϕ está contido no conjunto de crenças do agente i ou;

2. $\phi \in KB_i$; isto é, ϕ está contido no conjunto básico de crenças ou;

3. $KB_i \vdash_{p(i)} \phi$; ϕ é derivável do conjunto básico de crenças.

Outro detalhe importante a ser comentado diz respeito às propriedades das estruturas de dedução. Seguindo a estratégia de Konolige (1984, p. 23), apresentamos as propriedades e em seguida fazemos um breve comentário sobre cada uma delas.

PROPRIEDADE DA LINGUAGEM. A linguagem de uma estrutura de dedução é uma linguagem lógica. Essa é a única condição que Konolige impõe à linguagem: que essa linguagem seja uma linguagem lógica.

> Linguagens lógicas são conhecidas por terem um conjunto construtível de objetos sintáticos, as sentenças da linguagem, junto com algum método de interpretação[83]. (KONOLIGE, 1984, p. 21)

PROPRIEDADE DA DEDUÇÃO. As regras de uma estrutura de dedução são regras de dedução. O termo "regras de dedução" expressa certas propriedades importantes dessas regras, e as diferenciam de outros tipos de regras. As regras de dedução, de uma estrutura de dedução, possuem as seguintes propriedades:

1. As regras de dedução são todas "corretas"; ou seja, não é possível obter conclusões falsas a partir de premissas verdadeiras;
2. As regras de dedução são efetivamente computáveis;
3. As regras de dedução possuem um número finito de premissas.

PROPRIEDADE DE FECHO. O conjunto de crenças de uma estrutura de dedução é o menor conjunto que inclui as sentenças básicas, e é fechado sob deduções[84];

PROPRIEDADE DE RECURSÃO. Os agente vêem outros agentes como tendo um subconjunto de crenças similares aos seus próprios (KONOLIGE, 1984, p. 35). Ou seja, um modelo para sentenças envolvendo crença é, ele mesmo, um conjunto de crenças de uma estrutura de dedução.

Konolige (1984, p. 24-38) discute com detalhe cada uma das propriedades apresentadas, ora explicando-as, ora defendendo-se de possíveis críticas. Nesta seção, nos limitaremos a fazer pequenos comentários sobre elas.

[83]"*Logical languages are distinguished by having a constructable set of syntatic objects, the sentences of the language, together with an interpretation method.*"

[84]Isso não significa a mesma coisa que onisciência lógica; detalhes serão esclarecidos mais à frente.

Propriedade da linguagem

É de suma importância entender o motivo pelo qual a linguagem das estruturas de dedução seja uma linguagem lógica. Primeiramente, deve-se observar que se busca um modelo para as CRENÇAS. Ao se falar sobre as crenças de alguém (ou de algo, no caso da inteligência artificial), imediatamente somos levados a considerar conceitos como "verdade", "falsidade", etc; ou seja, é impossível falar sobre as crenças de alguém sem se remeter a algum tipo de semântica. Não devemos de modo algum confundir isso com as abordagens semânticas discutidas anteriormente. Nessas abordagens, os agentes manipulam proposições (que contêm, portanto, significado); enquanto que, na abordagem sentencial, os agentes lidam apenas com sentenças, uma lista de fórmulas. A introdução de uma semântica faz-se necessária, porque:

1. Se estamos a falar das crenças dos agentes, frequentemente vamos querer saber se tais crenças são ou não verdadeiras no mundo real;

2. Além disso, há também a preocupação com as regras utilizadas pelos agentes na derivação de suas crenças; isto é, constantemente nos perguntaremos se uma ou outra regra de dedução utilizada na derivação de uma crença é uma regra *correta*.

Para Konolige:

> Tais conceitos não fazem sentido na falta de um método de interpretação – um modo sistemático de interpretar as construções da linguagem em termos de um modelo. [...] Nós não podemos simplesmente colocar o referente de "Cícero" dentro de nossas cabeças, mesmo se ele estivesse vivo. Mas a atribuição de uma semântica para as sentenças é necessária, se um observador externo for analisar a natureza das crenças de um agente[85]. (KONOLIGE, 1984, p. 24)

A linguagem lógica possui, portanto uma semântica. Essa linguagem deve servir como um parâmetro para o modelo formal. Deste modo, para toda linguagem lógica $\mathscr{L}$, existe um conjunto de estruturas de dedução $D(\mathscr{L})$ das quais os conjuntos de sentenças básicas são sentenças da linguagem $\mathscr{L}$ (KONOLIGE, 1984, p. 26).

[85]*"Such concepts make no sense in the absence of an interpretation method – a systematic way of interpreting the constructions of the language in terms of a model. [...] We simply cannot put the referent of "Cicero" inside our heads, even if he were alive. But the attribution of semantics to sentences is necessary if an outside observer is to analyze the nature of an agents beliefs."*

Vale ressaltar, contudo, que o sistema dedutivo de crenças é independente da semântica; isto é, a princípio, como é comum, é apenas um conjunto de símbolos. Nenhum significado precisa ser dado às sentenças se não houver uma preocupação de dizer quando as crenças dos agentes são verdadeiras ou falsas; ou quando não houver interesse em saber se as regras que um agente utilizou em uma derivação estão corretas. O sistema dedutivo funciona por si só: é possível entender como o agente deriva as sentenças através da observação do conjunto básico de crenças e das regras de dedução. Porém, saber se aquilo que é derivado é "verdadeiro" ou não é um trabalho próprio da semântica.

Propriedade da dedução

Vale salientar novamente as propriedades importantes das regras de dedução:

1. Efetividade: uma regra de dedução é uma função efetivamente computável das sentenças da linguagem $\mathscr{L}$;
2. Provincialidade[86]: As regras de dedução devem possuir um número fixo e finito de premissas.
3. Correção: a conclusão é correta com relação à semântica.

Outras duas propriedades também são importantes para o sistema dedutivo:

4. Reflexividade: $\phi \vdash_{p(i)} \phi$;
5. Fecho: $\mathbf{B} \vdash_{p(i)} \phi$ e $\mathbf{B} \cup \{\phi\} \vdash_{p(i)} \psi$, então $\mathbf{B} \vdash_{p(i)} \psi$.

Pela provincialidade, podemos derivar o seguinte resultado:

Teorema 2.35. *Monotonicidade: Se S e S' são conjunto de sentenças, $S \subseteq S'$ e $S \vdash_{p(i)} \phi$, então $S' \vdash_{p(i)} \phi$. Ou seja, se ϕ é uma consequência de um conjunto de sentenças S, por meio de uma regra R qualquer, então ϕ também é consequência de qualquer conjunto maior $S' \supset S$.*

[86]*Provinciality.* (KONOLIGE, 1984, p. 26)

Demonstração. Se ϕ é derivável a partir do conjunto de sentenças S, então é possível construir uma prova:

$$
\begin{array}{r|l}
1 & S \\
\cline{2-2}
2 & . \\
3 & . \\
4 & . \\
5 & \phi
\end{array}
$$

Seja Γ o conjunto de sentenças S'. Se $S' \supset S$, então podemos construir uma prova de ϕ a partir de $S \cup \Gamma$, aplicando as mesmas regras de dedução em $S \cup \Gamma$:

$$
\begin{array}{r|l}
1 & S \\
2 & \Gamma \\
\cline{2-2}
3 & . \\
4 & . \\
5 & . \\
6 & \phi
\end{array}
$$

Logo, se $S \subseteq S'$ e $S \vdash_{p(i)} \phi$, então $S' \vdash_{p(i)} \phi$. □

Regras de inferência estendidas. O sistema dedutivo de crenças é basicamente formado pelas regras de dedução. Todas as regras de dedução possuem as propriedades que foram comentadas mais acima. Além das regras de dedução, há também as regras de inferência estendidas, que também são úteis na derivação de certas sentenças. Todavia, diferentemente das regras de dedução, as regras estendidas não satisfazem necessariamente todas as propriedades mencionadas. Konolige opta por não utilizá-las, em princípio, no sistema dedutivo de crenças (KONOLIGE, 1984, p. 31). A seguir, alguns exemplos de regras estendidas:

Revisão de crenças: o conjunto de crenças de um agente é constantemente atualizado, isso para estar sempre consistente com uma nova informação (de fato, pois, inevitavelmente, o agente entrará sempre em contato com novas informações).

"Raciocínio descuidado"[87]: o agente "pula para uma conclusão". Exemplo de "raciocínio descuidado" apresentado por Konolige:

> Se x é um pássaro, e nada do que se sabe de x o contradiga, assuma que x possa voar[88]. (KONOLIGE, 1984, p. 28)

Raciocínio introspectivo: o agente tira conclusões sobre o mundo baseado no conhecimento acerca de seu próprio conjunto de crenças. Exemplo de argumento com raciocínio introspectivo:

> Houve alguma presidente feminina? Não posso certamente nomear todos os presidentes. Por outro lado, não sei de qualquer presidente feminina. Se tivesse tido alguma, eu teria sabido. Logo, não deve ter havido qualquer presidente[89]. (KONOLIGE, 1984, p. 29)

As derivações baseadas na revisão de crenças não são consideradas por Konolige, por serem muito complicadas. Já a regra do "raciocínio descuidado" é observada mais a fundo. Primeiramente, pode-se observar que ela não tem a propriedade de ser monotônica: simplesmente adicione a sentença "x é um avestruz" àquele conjunto básico de crenças do exemplo. Agora, mesmo formando um conjunto de crenças básicas maior do que o anterior, o agente não consegue mais derivar "x pode voar". Isto é, mesmo admitindo $S' \supset S$ e $S \vdash_{p(i)} \phi$, não se obtém $S' \vdash_{p(i)} \phi$. Além de não-monotônica, a regra do "raciocínio descuidado" também não é correta, dado que é possível encontrar uma circunstância na qual x não possa voar, mesmo sendo um pássaro. Vale ressaltar também que a regra do "raciocínio descuidado" também não é provincial: é fácil observar que ela não possui um número fixo e finito de premissas. Konolige observa:

> [...] todas as regras de inferência não-monotônicas devem ser não-provinciais. Uma regra de raciocínio descuidado como a regra F especifica que algo deve ser consistente com um conjunto de sentenças antes que a regra possa ser aplicada[90]. Isso é o que a torna *não-provincial*: as premissas da regra de inferência não são um conjunto finito, fixo[91]. (KONOLIGE, 1984, p. 29)

[87] Aqui, devemos salientar que a utilização do termo "raciocínio descuidado" é de nossa inteira responsabilidade. O termo original é *default reasoning* que seria melhor traduzido por "raciocínio por falha".

[88] "*If x is a bird, and nothing that is known about x contradicts it, assume that x can fly.*"

[89] "*Were there any female presidents? I certainly can't name all the presidents. On the other hand, I don't know of any female presidents, and if there had been any, I would have known it; therefore there mustn't have been any.*"

[90] Assim é chamada a regra de raciocínio descuidado, exposta um pouco acima.

[91] "[...] *all nonmonotonic inference rules must be nonprovincial. A default rule like Rule F specifies that something must be consistent with a set of sentences before it can be applied. This is what makes it nonprovincial: the premisses of the inference rule are not a fixed, finite set.*"

Igualmente à regra do "raciocínio descuidado", a regra de introspecção também é não-provincial. O exemplo da feminista que se pergunta sobre a existência de um presidente do sexo feminino na história americana mostra claramente essa diferença. A inferência é feita na dependência do subsistema de crenças como um todo. Todavia, a regra de introspecção possui uma diferença com relação ao raciocínio descuidado. Claramente, é possível perceber que a inferência que a feminista faz é baseada em suas próprias crenças. A falta de uma sentença qualquer dentro do subsistema de crenças pode levar a feminista a inferir algo falso. Suponha que houve uma presidente feminina nos Estados Unidos, mas a feminista, por alguma razão particular, não foi informada[92]. Daí, a feminista falha na inferência pela falta de uma sentença particular no subsistema de crenças do tipo: "*a* foi presidente dos Estados Unidos", sendo *a* do sexo feminino.

Qual seria então a diferença entre a regra do raciocínio descuidado e a regra da introspecção? A resposta é que o raciocínio introspectivo é correto: as inferências que um agente faz sobre a falha de um conhecimento particular são justificadas, se a maneira de se obter as informações para a inferência estiver correta; isto é, se o agente estiver justificado ao fazer a inferência. A feminista falhou a saber da existência de um presidente do sexo feminino devido a forma insuficiente de obter a informação de que precisava. Porém, se a feminista estivesse de posse dessa informação, claramente concluiria que já existiu um presidente do sexo feminino.

Por não obedecer os mesmos critérios que as regras de dedução, as regras estendidas não fazem parte das estruturas de dedução.

Propriedade de fecho

O fecho das regras de inferência sob dedução é uma alternativa para "descomplicar" as coisas. Se não houvesse o fecho sob dedução das regras, teria de haver uma estratégia de controle complicada na derivação de crenças. Relembremos que é a estratégia de controle que garante como as regras de dedução serão aplicadas, e para onde irão as respostas feitas ao subsistema de crenças. A estratégia de controle também determina quais linhas de raciocínio seguir ou não. Modelar isso seria muito difícil; por conta disso, Konolige prefere assumir o fecho sob dedução das regras de inferência.

[92]Digamos que tenha havido alguma conspiração da CIA para encobrir a existência da uma presidente feminina.

A partir daí, uma pergunta importante pode ser feita: na abordagem sentencial, os agentes são logicamente oniscientes?

Não necessariamente. Para explicar esta resposta, é preciso fazer uma distinção entre *fecho sob dedução* e *fecho consequencial* (onisciência lógica). No entanto, antes de fazer tal distinção, convém, em primeiro lugar, apresentar algumas definições[93]:

Definição 2.19. *A linguagem $\mathscr{L}_D$ é uma linguagem padrão de primeira ordem acrescida do operador para crenças B_i, para cada agente i. As regras de formação para as fórmulas de $\mathscr{L}_D$ são dadas a partir da adição da seguinte cláusula a uma linguagem padrão $\mathscr{L}$ de primeira ordem:*

- *Se ϕ é uma sentenças de $\mathscr{L}$, então $B_i\phi$ é uma sentença de $\mathscr{L}_D$.*

Definição 2.20. *Uma estrutura de dedução para um agente i é um par $d_i = \langle p(i), \mathscr{L} \rangle$ no qual $p(i)$ é um conjunto de regras de dedução e $\mathscr{L}$ é a linguagem interna do agente. D é uma sequência de estruturas $\{d_o, d_1 ...\}$, uma para cada agente.*

Definição 2.21. *Uma estrutura modelo padrão de primeira ordem é uma estrutura $\langle \mathscr{U}, \mathscr{I}, \pi \rangle$, na qual $\mathscr{U}$ é o universo de indivíduos, $\mathscr{I}$ é uma função que vai de constantes de de $\mathscr{L}_D$ aos indivíduos e π é uma atribuição de verdade a todas as sentenças atômicas fechadas de $\mathscr{L}_D$*

Agora, apresentamos a distinção entre os dois tipos de fecho:

Fecho sob dedução. Toda sentença que pode ser derivada de um conjunto de sentenças KB_i, por meio das regras $p(i)$, está no conjunto de crenças da estrutura $d_i = \langle p(i), \mathscr{L} \rangle$. Isto é, o fecho sob deduções garante que todas sentenças do conjunto básico de crenças, juntamente com aquelas sentenças deriváveis a partir das regras de dedução, estejam no conjunto de crenças. Formalmente: Se $\mathbf{B} \vdash_{p(i)} \phi$ e $\mathbf{B} \cup \{\phi\} \vdash_{p(i)} \psi$, então $\mathbf{B} \vdash_{p(i)} \psi$.

Fecho consequencial. Toda consequência lógica está no conjunto **B**; isto é, toda consequência lógica está no conjunto de crenças. Isso é o mesmo que dizer que o agente acredita em todas as consequências lógicas de seu conjunto básico de crenças. Em outras palavras, o agente é logicamente onisciente.

[93]Na intenção de simplificarmos a apresentação da abordagem sentencial de Konolige (que é muito extensa), escolhemos apresentar as definições expostas em Whitsey (WHITSEY, 2003, p. 22-23).

A diferença entre os dois é clara. Se as regras de inferência são incompletas, haverá consequências do conjunto básico de crenças e das regras que não estarão no conjunto de crenças. Ou seja, se as regras de inferência são incompletas, o agente não precisa acreditar necessariamente em todas as consequências lógicas de seu conjunto básico de crenças. Isso porque haverá uma ou outra dedução que ele não conseguirá terminar, devido a suas limitações. Assim, os princípios (E-CLOS 1), (E-CLOS 2), (E-CLOS 3) e (E-CLOS 6) são todos passíveis de invalidação na abordagem epistêmica sentencial. Isso, contudo, não é afirmar que são inválidos. Apenas que, se for de interesse da lógica – e isso depende do que se pretende modelar – há meios formais para garantir que nenhum deles seja uma propriedade válida dentro dela.

Propriedade da recursão

O subsistema de crenças deve ter a capacidade de lidar com reiterações do operador de crença. Se o subsistema de crenças tem essa capacidade, diz-se então que ele tem a propriedade de recursão. Seu funcionamento é o seguinte: os agentes veem outros agentes como tendo um subsistema de crenças similares aos seus próprios. Isso não seria uma certa limitação do subsistema de crenças? Há aí algum problema de representação? Segundo Konolige:

> Isso ainda deixa uma grande flexibilidade para representar as crenças reiteradas. Por exemplo, John pode acreditar que a linguagem interna de Sue é $\mathscr{L}_1$, e que ela possui um conjunto de regras derivacionais $\mathscr{R}_1$, e que a linguagem interna de Kim é $\mathscr{L}_2$, e suas regras $\mathscr{R}_2$. Em adição, John pode acreditar que Sue acredita que a linguagem interna de Kim é $\mathscr{L}_3$, e suas regras $\mathscr{R}_3$[94]. (KONOLIGE, 1984, p. 35)

Já em um certo nível de reiteração, faz-se necessário especificar de quem é a estrutura de dedução. Como se faria, por exemplo, para especificar o subsistema do agente c pelo modo como o agente a acredita que o agente b o veja? Tome-se como exemplo os mesmos nomes utilizados por Konolige, a saber, John, Sue e Kim. Como fazer para representar o subsistema de crenças de Kim pelo modo como John acredita que Sue o veja? Isso é feito da seguinte maneira: $\nu = John, Sue, Kim$

Isto significa: "John acredita que Sue acredita que o subsistema de crenças

[94]*"This still leaves a large amount of flexibility in representing nested beliefs. For example, John might believe that Sue's internal language is $\mathscr{L}_1$ and that she has a set of derivational rules $\mathscr{R}_1$, whereas Kim's internal language is $\mathscr{L}_2$ and her derivational rules are $\mathscr{R}_2$. In addition, John might believe that Sue believes that Kim's internal language is $\mathscr{L}_3$, and her rules are $\mathscr{R}_3$."*

de Kim é...”. A letra grega v é utilizada para especificar de quem é o ponto de vista. A estrutura de dedução que formaliza esse ponto de vista é então $d_{John,Sue,Kim}$ (KONOLIGE, 1984, p. 36). O mundo real é representado pelo ponto de vista $v = \emptyset$. Um ponto de vista simples, envolvendo somente uma estrutura de dedução, é representado por $v = Kim$; uma estrutura de dedução para isso seria a estrutura d_{Kim}. A função $p(v)$ serve para especificar conjuntos de regras de dedução para cada ponto de vista v.

2.4.4 Alguns resultados

Antes de comentar alguns dos resultados do modelo dedutivo de crenças, faz-se necessário falar do termo “família de lógicas”. Com base nas estruturas de dedução, é possível construir toda uma família de lógicas; sistemas análogos ou não aos sistemas padrão **K**, **T**, **S4**, **S5**, etc. Cada sistema é considerado uma lógica particular; cada lógica da família $\mathbf{B}(\mathscr{L}, p)$ é uma axiomatização das estruturas de dedução $\mathscr{D}(\mathscr{L}, p)$. O símbolo para a linguagem de **B** é $\mathscr{L}_{\mathbf{B}}$. A linguagem $\mathscr{L}_{\mathbf{B}}$ inclui operadores modais, que servem para afirmar quando uma dada sentença é uma crença do agente. A linguagem $\mathscr{L}_{\mathbf{B}}$ não é a mesma linguagem do agente: $\mathscr{L}_{\mathbf{B}}$ pode ser entendida como a linguagem de um observador externo. (KONOLIGE, 1984, p. 39)

As derivações de teoremas são feitas com o método de tablôs de Hintikka, e as regras utilizadas nos tablôs são em forma de sequentes. Para detalhes, ver Konolige (1984). A seguir, uma definição importante.

Definição 2.22. *Um $\mathbf{B}(\mathscr{L}, p)$–modelo é uma quíntupla $\langle \mathscr{U}, \mathscr{I}, \pi, \mathscr{D}, \mu \rangle$, na qual $\mathscr{L}$ é a linguagem de todas as estruturas de dedução e as regras de d_i são $p(i)$.*

Como já foi comentado, o modelo dedutivo de crenças não gera, a princípio, onisciência lógica; isso porque as deduções podem ser incompletas, fazendo com que algumas consequências lógicas não estejam no conjunto de crenças. É possível interferir no conjunto de crenças dos agentes, fazendo restrições às regras que esses agentes manipulam. É possível, também, diminuir a quantidade de restrições; isso aumenta, por exemplo, a quantidade de regras que um agente manipula. Fazendo isso, pode-se considerar modelos particulares que irão corresponder àqueles da abordagem tradicional dos mundos possíveis. Konolige apresenta várias famílias lógicas, cada uma correspondendo a um sistema tradicional.

Por exemplo, a classe dos modelos B_K valida o esquema $B_i(p \to q) \to (B_i p \to B_i q)$, isto é, (E-CLOS 1). Os modelos B_K fazem parte de uma classe particular de modelos de **B**; essa é a classe de sistemas de crenças nos quais as regras de dedução são corretas e completas. Vimos que, no modelo dedutivo de crenças, o que impede de os agentes serem logicamente oniscientes é o fato de as regras de dedução serem incompletas. Daí, nos modelos B_K, como as regras são completas, é possível obter o resultado da onisciência lógica (dado que todas as consequências lógicas estarão no conjunto de crenças). Konolige chama esse sistema dedutivo de "família lógica não-introspectiva B_K"[95]. A classe dos modelos B_K validam, portanto, $B_i(p \to q) \to (B_i p \to B_i q)$; daí, segue-se que, em B_K, as crenças dos agentes são fechadas sob implicação.

Pode-se também construir as classes de modelos B_T B_4 e B_5, respectivamente. Esses modelos irão corresponder aos conhecidos sistemas de lógicas modais padrão **T**, **S4** e **S5**. Deste modo, os axiomas:

T $B_i p \to p$

S4 $B_i p \to B_i B_i p$

S5 $\neg B_i p \to B_i \neg B_i p$

são todos deriváveis no modelo dedutivo de crenças. É possível também construir classes de modelos que combinam essas propriedades. Isto é, podemos combinar B_K + B_4, e obter a classe de modelos B_{K4}, que corresponde ao sistema modal **K4**. Em suma, é possível obter todas as combinações de sistemas modais, assim como na abordagem padrão. Isso significa que, adaptações podem ser feitas de tal modo ao sistema dedutivo de crenças que todos os resultados da lógica epistêmica padrão podem ser deriváveis a partir do modelo sentencial, inclusive a onisciência lógica. Entretanto, o modelo sentencial pode igualmente ser construído de tal modo que o sistema não mantenha entre seus teoremas a propriedade de onisciência lógica. Alguns comentários devem ser feitos sobre isso em seguida.

2.4.5 Comentários sobre a abordagem sentencial

A abordagem sentencial permite invalidar os quatro casos de onisciência lógica. Isso pode ser feito através de manipulações nos subsistemas de crenças

[95] O termo "não introspectivo" é utilizado para mostrar que os esquemas introspectivos não são válidos nos modelos B_K.

dos agentes; basicamente, manipulando as regras que os mesmos conhecem e dominam. Deste modo, uma das formas de eliminar a onisciência lógica é reduzir a quantidade de regras que os agentes manipulam. Se uma dada regra é necessária à derivação de uma consequência lógica, e não sendo permitido a um agente utilizar essa regra, então o agente falha em derivar o resultado. Vejamos, a seguir, algumas das limitações que podem ser modeladas:

(1) Recursos limitados de tempo. No sistema dedutivo de crenças é possível também reproduzir a falha de onisciência lógica por falta de tempo. Isso acontece ao se alterar, por exemplo, algumas regras de dedução. Observe-se o exemplo a seguir. Suponha que um agente qualquer, digamos, a, que usa a regra *modus ponens* na derivação de crenças. Ou seja, suponha que o agente manipula a regra $\frac{p, p \rightarrow q}{q}$ (MP). Esta regra poderia ser alterada de modo que a limitação de tempo fosse satisfeita. A regra resultante seria mais ou menos assim:

$$\frac{DD(n) \wedge p, DD(m) \wedge (p \rightarrow q)}{DD(n+m+1) \wedge q} \ (MP),\ n+m \leq k.$$

'*DD*' é o chamado "custo derivacional". A regra, portanto, soma o custo derivacional de p e $p \rightarrow q$ + 1, que é o custo da inferência de q. A letra k representa um número natural qualquer; k é o limitador de tempo. Para modelar a limitação de tempo, basta dar um número a k, n e m. Isso significa o mesmo que estipular ao agente, para a obtenção de q, um número finito de passos. A inferência de q só poderá ser obtida se a soma de $n+m+1$ for menor ou igual a k. Se essa soma for maior, então o agente não conseguirá derivar q. Assim, a limitação de tempo é representada pelo número finito de passos que o agente pode dar em uma derivação. Ou seja, a limitação de tempo é representada de um modo inteiramente sintático.

(2) Ausência de consciência de regras relevantes. Isso já foi explicado mais acima. A restrição ao uso de regras de inferência que o agente manipula interfere diretamente na capacidade do agente de derivar consequências lógicas do seu conhecimento.

Ao que nos parece, os outros motivos não são satisfeitos pelo modelo dedutivo. Como era de se esperar, assim como as outras abordagens, o modelo dedutivo satisfaz parcialmente os motivos de falha de onisciência lógica apresentados no capítulo 1 e neste capítulo. As limitações do modelo dedutivo são evidentes:

- Reduzir o processo de obtenção de crença (ou conhecimento) de agentes a derivações lógicas não dá conta da complexidade do tema. O processo de obtenção de conhecimento e crenças dos seres humanos comportam fatores que estão além do escopo da lógica.

Devemos, contudo, relembrar que o modelo dedutivo proposto por Konolige é direcionado à inteligência artificial. É claro que certas semelhanças surgem, quando se compara, por exemplo, um subsistema de crença ao raciocínio humano. Porém, como era de se esperar, as enormes diferenças entre agentes humanos e máquinas (ou programas) impedem que o modelo dedutivo de crenças represente completamente as falhas e capacidades do raciocínio humano.

Em suma, o conhecimento e crenças dos seres humanos é obtido com algo a mais do que um processo dedutivo, puramente sintático. De qualquer modo, não tiramos de modo algum a utilidade desta abordagem. Assim como as outras abordagens, o modelo dedutivo de crenças satisfaz aspectos particulares da lógica epistêmica; lida com a obtenção de crenças a partir de processos dedutivos. Apesar da não constituir em totalidade o modo como pensamos e obtemos nossas crenças, a derivação lógica certamente ocupa um lugar importante e indispensável nesse processo. Novamente, aqui, sugiro que a intenção de aplicação do nosso modelo epistêmico (ou epistemológico) nos dirá que aspectos deverão ser considerados, além das especificidades dos contextos e dos agentes que serão modelados, bem como o escopo do conceito de validade.

2.5 Onisciência lógica: problema solucionado?

Pudemos observar que todas as abordagens apresentadas ofereceram soluções para problema da onisciência lógica. Em certos momentos, elucidamos pontos positivos de cada abordagem; em outros, os pontos negativos também foram considerados. Não se pode negar que cada solução propôs um método criativo para solucionar o problema. A pergunta que fazemos, então, é a seguinte:

Das soluções que foram apresentadas, alguma foi bem sucedida em solucionar o problema da onisciência lógica?

Ora, pelo menos em princípio, conseguimos pensar em duas respostas:

1. Todas elas solucionaram o problema.

2. Nenhuma delas solucionou o problema.

Qual das duas respostas é a melhor, a primeira ou a segunda? Na verdade, as duas podem ou, talvez, nenhuma delas; isso vai depender do modo como vemos o problema, e o que consideramos como “solução” para esse problema.

2.5.1 Discussão sobre a resposta 1

Aqui, devemos começar com uma outra questão: sob quais circunstâncias as abordagens apresentadas solucionariam o problema da onisciência lógica? Para responder esta questão, por sua vez, pensemos ainda em uma outra: é necessário, a uma lógica epistêmica, satisfazer todos os motivos pelos quais um agente falha em ser logicamente onisciente? Ou então: as lógicas epistêmicas devem satisfazer todos os motivos de falha de onisciência lógica apresentados neste e no primeiro capítulo? Ora, durante a exposição das abordagens, ficou claro que nossa resposta a estas duas últimas perguntas é: não necessariamente. Vejamos o porquê.

Entre os capítulos 1 e 2, explicitamos alguns dos motivos pelos quais os agentes falham em ser logicamente oniscientes. Isto é, fizemos distinção entre pelo menos quatro motivos pelos quais um agente com capacidades racionais limitadas não consegue derivar todas as consequências lógicas daquilo que ele conhece ou acredita:

1. Recursos computacionais limitados (*resource-bounded*). Nem sempre conhecemos as regras relevantes para derivação de certas consequências lógicas daquilo que conhecemos; ou então, mesmo as conhecendo, nem sempre dispomos de tempo suficiente para realizar tal derivação.

2. Ausência de consciência dos conceitos relevantes (*lack of awarenes)*;

3. Preconceito. Um agente pode falhar em encontrar certas consequências lógicas de seu conhecimento devido a certos preconceitos que possui.

4. Atenção desconexa. As pessoas não conseguem prestar atenção a todos os temas simultaneamente.

O que devemos nos perguntar é se, em nossa vida cotidiana, falhamos em ser logicamente oniscientes devido a todas essas razões ao mesmo tempo. Obviamente, a resposta é não. Ora falhamos em uma, ora em outra; ora falhamos ainda em alguma outra, não listada entre as razões que apresentamos aqui. Em suma, é razoável pensar ser improvável que falhemos em todas elas ao mesmo tempo.

Para ilustrar essa ideia, suponha, por exemplo, um estudante de filosofia que não aprendeu bem os métodos de dedução natural ensinados na disciplina de lógica elementar. Suponha que é dada a ele a seguinte tarefa:

- Escolhe-se um argumento válido qualquer. Pede-se ao estudante que faça uma prova lógica do argumento, utilizando o método de dedução natural. São dadas a ele uma lista de premissas, as quais já estão organizadas no topo de uma folha de papel, e a conclusão do argumento (que já está localizada na parte inferior do mesmo papel). Suponha que, se efetuada corretamente, a prova do argumento requer apenas dez linhas. Imaginemos que o argumento termina com conclusão de que Sócrates foi um filósofo. Suponha também que todos os conceitos envolvidos na prova são conhecidos pelo estudante (ex. filosofia, filósofo, etc.). Ora, se isso é o caso, podemos dizer que o estudante está ciente de todos os conceitos relevantes. Deste modo, neste caso, um motivo de falha de onisciência lógica já não precisa ser satisfeito (falta de ciência de conceitos relevantes à derivação). Como já foi dito, o estudante não domina o método de dedução natural. Assim, apesar de a sentença "Sócrates foi um filósofo" ser consequência do que o estudante conhece, ele não é capaz de realizar a prova. O agente falha na obtenção da prova devido aos recursos limitados que possui em dedução natural – isto é, falta de perícia na manipulação de regras lógicas relevantes à derivação. Portanto, apesar de estar ciente dos conceitos relevantes, o agente está inapto a realizar a prova do argumento.

O que esse exemplo nos mostrou? Que o referido agente não conseguiu realizar a prova por satisfazer o motivo (2) de falha de onisciência, mas não o motivo (1). Isto é, neste caso, nem todos os motivos de falha de onisciência lógica foram necessários para que o agente falhasse em ser logicamente onisciente.

Agora imagine que, além de o agente estar ciente dos conceitos relevantes, de possuir recursos computacionais e de tempo suficientes e de pensar de modo coerente – isto é, sem considerar impossíveis mundos possíveis – o agente pode falhar em obter a prova do argumento por satisfazer o motivo número (3). Isto é, apesar de possuir todos os recursos necessários à obtenção da prova do argumento, o agente não o faz devido a certos preconceitos, ou simplesmente por não se interessar na tarefa. Deste modo, novamente neste caso, necessitamos apenas de um motivo particular (e não todos) para satisfazer a falha em onisciência lógica.

Podemos mostrar exemplos similares com todas as combinações entre os cinco motivos apresentados, e mostrar que sempre é possível que um deles não seja satisfeito. Deste modo, não se faz necessário que, para toda lógica epistêmica, os cinco motivos de falha de onisciência lógica devam ser satisfeitos ao mesmo tempo. Logo, se uma dada lógica satisfaz ao menos um motivo desses, ela já é útil para modelar um aspecto da limitação lógica dos agentes (sejam eles artificiais ou humanos).

Logo, nesta perspectiva, cada abordagem apresentada resolve o problema da onisciência lógica. As circunstâncias nas quais cada abordagem oferece uma solução ao problema são aquelas em que o agente satisfaz um ou outro motivo particular de falha em onisciência lógica. Essas abordagens são, portanto, soluções que se aplicam a situações particulares. Agora, se é requerido que, para cada uma delas, todos os motivos da falha em onisciência lógica sejam satisfeitos, então nenhuma das abordagens resolve o problema. Entramos, daí, na resposta número 2.

2.5.2 Discussão sobre a resposta 2

Em que circunstância nenhuma das abordagens apresentadas seria solução suficiente para o problema da onisciência lógica? A resposta dessa pergunta é fundamental para se compreender a perspectiva de que nenhuma das abordagens apresentadas representam a solução adequada.

As circunstâncias nas quais as abordagens se mostram insuficientes são aquelas em que é requerido às mesmas que satisfaçam motivos não-lógicos de falha de onisciência lógica. Explicaremos isso a seguir.

É interessante observar que o próprio termo “onisciência lógica” parece colocar o problema em uma caracterização puramente lógica, o que não deve ser o caso. É verdade que o problema se refere explicitamente às nossas capacidades lógicas. Porém, foi possível observar que nossas limitações em perseguir consequências lógicas são causadas não somente por fatores lógicos, mas também por conta de fatores externos ou psicológicos. Para esclarecer isso, observe-se o seguinte:

- Preconceito ou falta de interesse: um agente falha em derivar uma certa lógica por preconceito ou por falta de interesse.

- Limitação de tempo; um agente falha em derivar uma certa consequência lógica por falta de tempo.

Pode-se perguntar se algumas destas causas para a falha de onisciência lógica são causas lógicas. O que dizer? Certamente, há mais do que puramente "lógica" envolvido nisso.

Vejamos o caso da falta de interesse. Nem sempre agimos seguindo estritamente a razão. Isto é, a razão não é causa suficiente na derivação de informações. Um indivíduo desinteressado pode não computar consequências lógicas de seu conhecimento, mesmo que possua todos os recursos necessários para fazê-lo. Para ilustrar isso, suponha por exemplo um estudante de matemática com a tarefa de demonstrar um teorema qualquer. Suponha que o estudante possui todos os recursos necessários à obtenção da prova (domínio na manipulação de regras e definições, conhecimento dos conceitos relevantes etc). Porém, suponha que esse estudante é um aluno frustrado, e que há pouco tempo se deu conta de que queria ser um advogado. Deste modo, pouco lhe interessa o estudo da matemática e seus teoremas. Logo, o estudante falha na obtenção da prova porque não tem interesse suficiente para seguir todos os passos da derivação.

A falha do estudante deu-se por razões lógicas? Argumentamos que não. É claro que poderíamos procurar uma série de razões lógicas para a atitude do estudante mas, no final das contas, só o estudante saberia com precisão o porquê da desistência de encontrar esse resultado matemático. Os motivos poderiam ser vários, até mesmo lógicos. Mas isso não é garantido. A falha na obtenção da prova poderia ser atribuída a outros motivos.

O que dizer agora da falha em onisciência lógica por recursos limitados de tempo? Suponha, por exemplo, um cientista brasileiro que está prestes a fazer uma descoberta revolucionária: a cura para todos os tipos de câncer. Suponha que, continuando no ritmo em que se encontra, a pesquisa estaria concluída em alguns meses. Porém, infelizmente, poucos dias antes de chegar ao resultado esperado, o cientista é vítima fatal de uma "bala perdida", em um semáforo. O cientista falhou em derivar a informação por falta de tempo, e essa falha deu-se por fatores completamente externos a ele, independentes. Casos como esses não são contemplados totalmente por uma abordagem lógica. Assim, a falha em onisciência lógica por recursos limitados de tempo não deve ser o subsídio teórico principal ao se construir uma lógica epistêmica. O argumento a seguir mostra o porquê.

Argumento do assassino 47

O agente 47, também conhecido simplesmente como "Hitman", representa a ideia do assassino de aluguel ideal. Consiste de um clone humano, criado e treinado desde a infância por uma seita banida da igreja, com um único propósito: matar "todos os representantes do mal que andam sobre a terra". Toda sua educação foi preparada de modo que ele se tornasse o assassino perfeito: preciso e silencioso. Deste modo, o agente 47 é uma arma perfeita, que nunca falha em suas missões.

Imagine agora um professor universitário qualquer, que trabalha também na pesquisa de armas de destruição em massa. Como todo professor, ele está acostumado a responder perguntas em sala de aula. Ora, é natural que ele nem sempre saiba as respostas imediatamente. Em casos como esses, pede ao aluno para esperar até a próxima aula, para que possa pesquisar, pensar a respeito (fazer inferências lógicas) ou simplesmente se lembrar da resposta. Pensemos agora na seguinte história:

- Em uma certa segunda-feira, quando o professor dava aula, um aluno fez uma pergunta para a qual o professor não tinha uma resposta imediata. Como de costume, o professor pediu ao aluno que esperasse até a quarta-feira (dia da próxima aula). Porém, o professor mal sabia que o agente 47 havia sido contratado para matá-lo, devido a suas pesquisas em armas nucleares. O agente 47 escolheu a terça-feira como o dia ideal para cumprir a missão. Como ele nunca erra, daí... O professor foi morto na terça-feira, exatamente como o agente 47 havia planejado. Pobre do aluno, ficou sem o professor e sem a resposta esperada.

Com base em tudo o que discutimos sobre onisciência lógica, nos perguntamos:

- O professor sabia a resposta?

A princípio, somos levados a responder que sim. Todo esse processo de pergunta e resposta na aula seguinte é algo bem comum para todo professor. Geralmente, e principalmente quando a resposta já é uma consequência lógica do que já é conhecido, é esperado que o professor esteja de posse da resposta na próxima aula. Mas nesse caso, nunca será possível saber, com cem por cento de certeza, que o professor sabia de fato a resposta.

O que decorre, então, desse exemplo? O seguinte: nunca se sabe se uma dada consequência lógica de tudo aquilo que um agente sabe chegará a ser conhecida pelo agente; pois nunca se sabe quanto tempo lhe resta para derivar todos os resultados. Isto é, nunca se sabe se o agente 47 foi contratado para matá-lo. Ora, se nunca sabemos quanto tempo temos para derivar uma certa consequência lógica, então modelar a falha de onisciência lógica por recursos limitados de tempo passa a ser irrelevante para uma teoria lógica. Isso porque o argumento do assassino 47 trivializa a falta de tempo, tornando-a um fator completamente imprevisível. Porém, este resultado não precisa ser tão pessimista. O argumento do assassino 47 mostra-nos apenas que, em muitos casos, a falha de onisciência lógica se dá por motivos externos – isto é, não lógicos. Por conta disso, pode acontecer de uma lógica epistêmica nem sempre se adaptar a tais contingências e, portanto, deixar de modelar alguns aspectos relevantes da falha de onisciência lógica.

O "recurso limitado de tempo" é um motivo não-lógico para a falha na onisciência lógica, um motivo externo e independente aos agentes. Logo, não deve ser um dos subsídios principais na construção de uma teoria lógica do conhecimento e da crença – isto é, se tal teoria for puramente formal. Apesar disso, há certos casos em que uma abordagem lógica é capaz de modelar os recursos limitados de tempo. Isso vai ocorrer exatamente quando dois requisitos forem satisfeitos:

1. Seja garantido que não haja qualquer interferência externa no processo de derivação[96];
2. Estipular um prazo limite para a derivação; isto é, estabelecer que o processo de derivação seja finito, e indicar o número máximo de passos permitido para a derivação.

Assim, com algumas ressalvas, o modelo lógico é capaz de satisfazer a falha de onisciência lógica por recursos limitados de tempo (sempre tendo em mente que esse não deve ser o subsídio teórico principal, dado que seria impossível modelar os

[96]Isso seria o mesmo que obter do assassino 47 a garantia de que não iria mais matar o professor.

recursos limitados de tempo de uma forma completa, como mostra o argumento do assassino 47).

2.5.3 Conciliando as respostas 1 e 2

Será que as respostas 1 e 2 são totalmente incompatíveis? Ou será que a pergunta do início deste capítulo deveria ser melhor elaborada? Uma coisa que a resposta 2 nos ensinou é que há motivos de falha de onisciência lógica que não são de ordem puramente lógica; e daí, que é impossível para as abordagens discutidas nesse trabalho (que são todas lógicas) solucionarem o problema de modo definitivo. Todavia, a resposta 1 nos mostrou que nem sempre todos esses motivos de falha de onisciência lógica são satisfeitos ao mesmo tempo; e que, como foi visto, cada abordagem discutida nesse trabalho é capaz de lidar com o problema de forma particular – isto é, estudando um motivo em particular. Assim, ao menos no que se diz respeito ao plano lógico, cada abordagem discutida nesse trabalho tem uma estratégia de solução para o problema.

O fato é que nossa limitação em saber de todas as consequências lógicas do nosso conhecimento se deve a vários fatores. Às vezes lógicos, às vezes externos, às vezes psicológicos etc. Penso que as abordagens lógicas devam voltar sua atenção aos motivos lógicos de falha de onisciência lógica. A tarefa passa então a ser a identificação desses motivos lógicos, e o desenvolvimento de lógicas que satisfaçam esses motivos. Certamente, os motivos de falha de onisciência lógica apresentados neste capítulo não esgotam a lista de todos os motivos possíveis.

Como conciliar as duas respostas? Primeiramente, admitindo que as abordagens discutidas não satisfazem todos os motivos. De fato, porque há alguns que nem mesmo devem ser discutidos em um âmbito lógico. Em seguida, observar que todas as abordagens lógicas discutidas nesse trabalho lidam satisfatoriamente com motivos que podem ser discutidos em âmbito lógico, como:

- Consideração de "impossíveis mundos possíveis"; isto é, concepção de situações lógicas incoerentes.
- recursos computacionais limitados;
- desconhecimento de regras de dedução relevantes;
- falha em estar ciente de conceitos relevantes.

Assim, as abordagens discutidas neste capítulo não explicam, utilizando apenas estratégias lógicas, todas as causas pelas quais um agente racionalmente limitado falha em ser logicamente onisciente. Entretanto, todas eles são instrumentos que explicam a limitação lógica dos agentes quando os motivos dessa limitação puderem ser apresentados em termos lógicos.

Logo, concordamos que nem todas as abordagens solucionam o problema em sua totalidade; mas, ao mesmo tempo, todas oferecem soluções satisfatórias com relação aos aspectos lógicos desse problema – em aplicações específicas, segundo seus interesses de modelagem.

2.5.4 Uma breve discussão sobre as abordagens consideradas neste capítulo

No decorrer deste capítulo, foram apresentadas diversas soluções lógicas para o problema da onisciência lógica. Cada uma delas tem suas características próprias e seu modo peculiar de lidar com o problema. Cada uma satisfaz um ou outro motivo de falha de onisciência lógica. Cada uma delas elimina um ou outro caso de onisciência lógica. Uma pergunta inevitável que se faz sobre isso é se, dentre as abordagens apresentadas, alguma delas é superior às demais. Isto é, a seguinte questão pode ser suscitada:

Existem critérios suficientes capazes de decidir quando uma lógica epistêmica é mais adequada que outra para solucionar o problema da onisciência lógica, em um âmbito geral?

Deixamos claro, no início deste capítulo, que esse não é o tema central de nossa investigação e que, portanto não será pesquisado a fundo. Todavia, achamos conveniente expor o que há na literatura sobre essa questão. Whitsey escreve (2003, p. 6) que vários critérios foram propostos para se avaliar e comparar lógicas epistêmicas. Entre esses critérios, podemos destacar:

1. As crenças (conhecimento) não deveriam ser fechadas sob implicação;
2. os agentes não deveriam conhecer todas as tautologias;
3. a lógica deveria permitir agentes múltiplos;
4. a lógica deveria permitir reiteração de modalidades;

5. os agentes não deveriam considerar impossíveis mundos possíveis.

Poderíamos, deste modo, comparar as abordagens apresentadas verificando qual delas satisfaz a maior parte desses critérios[97]. No entanto, não faremos isso, pois adotamos a postura de que uma dada lógica pode satisfazer apenas um critério particular, para investigar um caso particular de falha na onisciência lógica[98]. Somos, portanto, a favor da idealização. Nessa perspectiva, uma lógica epistêmica não precisa reproduzir, fielmente, todas as falhas e capacidades racionais de agentes reais (tarefa que, ao que parece, lhe seria impossível). Entendemos a lógica epistêmica, portanto, como uma teoria com seus próprios interesses, livre até mesmo para pesquisar situações ideais nas quais os agentes gozam de propriedades como, por exemplo, a onisciência lógica. Esta posição parece assumir, implicitamente, uma certa perspectiva acerca da atribuição, de um modo geral, de uma lógica epistêmica. Que perspectiva seria essa? Vejamos a próxima seção.

2.5.5 As atribuições de uma teoria lógica epistêmica

Começamos esta seção com a seguinte questão:

O que é uma teoria lógica epistêmica?

Grosso modo, poderíamos responder que consiste de um "estudo lógico de noções como conhecimento e crença". Certamente, isso faria parte de uma resposta satisfatória. No entanto, ainda não é suficiente para esclarecer o que é a lógica epistêmica.

Ao que parece, a definição do que é a lógica epistêmica está fortemente ligada à sua finalidade. A lógica epistêmica, no princípio, não tinha a função de entender o que é o conhecimento ou a crença. Isso, como se sabe, é uma tarefa historicamente dada à epistemologia, não à lógica. Entretanto, Hendricks (2006) defende a tese de que a lógica epistêmica – ou, como prefere o referido autor, "epistemologia formal" – é uma teoria epistemológica por si só, mas que deveria sim

[97]Ainda assim, essa não é uma tarefa tão fácil. Como aponta Hadley (1991, p. 56), os pesquisadores em inteligência artificial não esclarecem quando uma dada teoria ou formalismo lógico deva ser considerado um modelo cognitivo. Além disso, existem usos diferentes para a noção de crença.

[98]O importante, portanto, é verificar que cada uma das abordagens discutidas satisfaz ao menos um dos critérios acima. Ex. A abordagem original de Hintikka satisfaz o critério 4; isso lhe possibilita investigar a introspecção positiva, negativa, transmissão de conhecimento etc. Já na Lógica de Levesque, o critério 1 é satisfeito, e assim por diante.

trabalhar juntamente com a epistemologia informal. Para Hendricks, o que a epistemologia formal tem a oferecer à epistemologia *mainstream* é muito mais do que uma simples modelagem lógica de algumas das noções mais importantes desta última, tais como "conhecimento", "crença", "consciência" etc. No entanto, apesar de ser bastante popular atualmente, além de promissora, essa é uma ideia relativamente recente, ainda a ser bastante debatida nos dois âmbitos: na epistemologia informal e também entre os lógicos.

Em *Knowledge and Belief*, Hintikka (1962, p. 3) esclarece que a meta principal de sua obra consiste em formular e defender critérios explícitos de consistência para certos conjuntos de sentenças – no caso, sentenças epistêmicas. Ou seja, o modelo epistêmico de Hintikka é um estudo puramente lógico. A partir disso, poderíamos pensar que não deveríamos esperar que uma lógica epistêmica explicasse no que consiste o conhecimento e a crença, mas apenas as circunstâncias sob as quais argumentos que comportam sentenças epistêmicas seriam válidos ou inválidos. No entanto, como era de se esperar de um campo de estudo tão vasto, o crescente desenvolvimento da lógica epistêmica acabaria por elevá-la da condição de mero "suporte" de teorias epistemológicas ao status de parceira na elucidação de conceitos de interesse típicos da epistemologia *mainstream*. Isso, como vemos, parece reforçar a posição de Hendricks de que a lógica epistêmica, ou epistemologia formal, é uma teoria epistemológica por si só.

Atualmente, o conhecimento de epistemologia formal se mostra cada vez mais necessário ao debate epistemológico informal. Os problemas do fecho epistêmico e da onisciência lógica são fortes evidências para essa afirmação, na medida em que percebemos, agora, que o problema do fecho epistêmico vai muito além de perguntar se um determinado princípio de fecho é válido ou não. Ou seja, o tratamento dos princípios de fecho segundo a dicotomia validade/invalidade é uma simplificação exagerada do problema da análise lógico-epistemológica dos princípios de fecho epistêmico. É justamente contra isso que tenho argumentado até agora, e tenho utilizado evidências da lógica epistêmica para sustentar essa posição. Ora, o próprio fato de utilizar exemplos da epistemologia formal (o problema da onisciência lógica) para fundamentar uma posição epistemológica informal (sobre a não-simplificação da análise lógico-epistemológica de princípios de fecho) demonstra o quão importante pode ser a lógica epistêmica na elucidação de questões da epistemologia *mainstream*. Entretanto, essa não é minha meta principal neste capítulo. Aqui, contento-me em mostrar as estratégias da lógica epistêmica para a solução do problema da onisciência lógica, e sugerir a utilização de estratégia similar na análise epistemológica de princípios de fecho epistêmico.

Por outro lado, não podemos negar, uma lógica epistêmica também pode ser entendida como um modelo lógico para uma teoria epistemológica. Essa não seria a perspectiva sugerida por esta investigação, na medida que esta última concebe a epistemologia formal como o conjunto de teorias formais sobre o conhecimento e a crença, e não meramente como o conjunto de modelos lógicos de suporte a teorias epistemológicas. Porém, mesmo se assim o fosse, ainda assim a importância da lógica epistêmica para a epistemologia ficaria demonstrada. Nessa perspectiva, a teoria epistemológica seria aquela responsável pela interpretação das noções de conhecimento e crença; a teoria lógica estudaria os critérios de consistência para essas noções, tudo isso visando um modelo lógico que seja compatível com a teoria epistemológica. Quando então, teríamos problemas? Quando um dado resultado do modelo lógico (isto é, um teorema) não for válido na teoria epistemológica, e vice-versa. Ex:

- Suponha que, em minha teoria epistemológica, a introspecção positiva é um princípio válido. Isto é, sempre que alguém sabe de algo, essa pessoa sabe que sabe. Agora suponha que eu construo uma lógica epistêmica na qual a fórmula $K_a p \rightarrow K_a K_a p$ é inválida. Ora, se traduzirmos a formula $K_a p$ como significando "o agente a conhece p", nosso resultado no modelo lógico é incompatível com aquele defendido pela teoria epistemológica. Nesse caso, se sustentamos que o princípio da introspecção positiva é correto, concluímos que o problema está com nosso modelo lógico, que gerou um resultado incompatível com aquele que defendemos.

Nessa perspectiva, a onisciência lógica surge como uma incompatibilidade entre modelo lógico e teoria epistemológica. Tratamos dessa questão a seguir.

2.5.6 Onisciência lógica: problema para quem?

Caso consideremos o problema da onisciência lógica como uma espécie de incompatibilidade entre modelo lógico e aquilo que é aceito por uma teoria epistemológica, devemos tentar identificar a origem dessa incompatibilidade. Como já foi visto, um sistema epistêmico pode distinguir vários casos de onisciência lógica. A seguir, listamos três deles:

1. Fecho sob implicação (E-CLOS 1): $K_a(p \to q) \to (K_a p \to K_a q)$;
2. Fecho sob implicação válida (E-CLOS 6): se $K_a p$ e $\vDash p \to q$, então $K_a q$;
3. Conhecimento de fórmulas válidas (E CLOS 3): se $\vDash p$, então $\vDash K_a q$.

Ao observarmos essas três propriedades mais de perto e o comportamento semântico dos operadores 'K' e '$\to$' dentro da lógica epistêmica, percebemos a causa do problema. As três propriedades de onisciência lógica são obtidas a partir da interpretação semântica desses operadores dentro do sistema epistêmico. Todavia, note-se que o axioma $\Box(p \to q) \to (\Box p \to \Box q)$, também conhecido como "distribuição de $\Box$ sob a implicação", é amplamente aceito em lógica modal alética; isto é, ele é válido em todos os sistemas padrão de lógica modal alética. Nessa perspectiva, parece, portanto, que o problema surge a partir da interpretação do operador K:

Um agente a conhece p se, e somente se, p é verdadeiro em todas as alternativas epistêmicas concebidas por a.

Eis a questão: a verdade de p em todos os mundos possíveis alternativos ao atual garante o conhecimento de p por a? A resposta é clara: não! Tal como mostrou Gettier (1963, p. 121-123), a verdade não é critério suficiente para se determinar o conhecimento de alguém sobre uma proposição qualquer. Certamente, se alguém conhece algo, esse algo deve ser verdadeiro. Mas ser verdadeiro não é suficiente, há algo mais, algo que não é captado pela interpretação oferecida para o operador K na lógica epistêmica. Note-se que, na interpretação acima, usa-se "se, e somente se,". Porém, o condicional "se p é verdadeiro em todas as alternativas epistêmicas concebidas por a, então a conhece p" é terminantemente proibido pelo argumento de Gettier.

O que deve ser observado, porém, é que os problemas suscitados por Gettier são posteriores à abordagem fornecida por Hintikka, em 1962. Deste modo, a lógica epistêmica de 1962 foi construída com base em uma interpretação insuficiente da noção de conhecimento; isto é, uma interpretação que mais tarde seria questionada por Gettier em uma plano epistemológico, e não lógico. Daí, parece-nos que o problema da onisciência lógica surge a partir de uma interpretação – num plano epistemológico – insuficiente para a noção de conhecimento. Essa interpretação foi, deste modo, captada pela lógica epistêmica, e pode ter sido a geradora do problema da onisciência lógica.

O que podemos concluir disso é que o problema da onisciência lógica pode não ser necessariamente de ordem puramente lógica. Pois, se a interpretação dada ao operador K em um sistema epistêmico for compatível com a noção epistemológica que temos acerca do ato de conhecer, então certamente o problema da onisciência lógica não seria de responsabilidade da lógica epistêmica que, nesta ocasião específica, foi concebida apenas como um modelo lógico para uma teoria epistemológica (que no caso em questão seria a legítima responsável pela interpretação da noção de conhecimento). Não devemos, no entanto, confundir esse argumento e inferir que problemas como o da onisciência são sempre de responsabilidade de uma teoria epistemológica. O argumento afirma, na verdade, que podem existir várias causas para problemas como esses, e entre elas:

1. Incompatibilidade entre modelo lógico e teoria epistemológica. Nesse caso, o modelo lógico gera resultados incompatíveis com aqueles defendidos pela teoria epistemológica. Ex: A onisciência lógica é uma propriedade inaceitável na teoria epistemológica, mas a lógica epistêmica valida:

   (a) $K_a(p \to q) \to (K_a p \to K_a q)$;

   (b) Se $K_a p$ e $\vDash p \to q$, então $K_a q$;

   (c) Se $\vDash p$, então $\vDash K_a p$.

   Assim, nesse caso, o problema da onisciência lógica é de responsabilidade da lógica epistêmica. O problema pode estar na interpretação do operador K dentro do sistema; a interpretação não traduziu fielmente o significado que a teoria epistemológica dá à noção de conhecimento.

2. Compatibilidade entre modelo lógico e teoria epistemológica, porém interpretação inadequada da noção de conhecimento num âmbito epistemológico. Nesse caso, é construído um modelo lógico no qual todos os resultados são compatíveis com aqueles obtidos pela teoria epistemológica. No entanto, entre os resultados obtidos estão também (a), (b) e (c). Análise posterior mostra que os postulados da teoria epistemológica levam aos mesmos resultados, em âmbito informal.

3. Compatibilidade entre modelo lógico e teoria epistemológica, e resultados "sustentáveis" por parte da teoria epistemológica. Nesse caso, além da compatibilidade de resultados entre a lógica epistêmica e a teoria epistemológica, resultados como (a), (b) e (c) não são gerados em quaisquer das duas.

Portanto, as causas de problemas como onisciência lógica, introspecção positiva, introspecção negativa, entre outros, podem variar. Também, se concebida como mero modelo de suporte lógico para uma teoria epistemológica, a lógica epistêmica nem sempre será responsável por problemas que sejam evidenciados pela formalização; se a lógica epistêmica for concebida para gerar um modelo compatível com uma teoria epistemológica, os problemas que surgirem a partir da interpretação da noção de conhecimento (crença) serão de responsabilidade desta última, que foi a fornecedora da respectiva interpretação.

2.5.7 A onisciência lógica é um problema?

Como já foi visto, há pelo menos duas perspectivas teóricas para se conceber uma lógica epistêmica:

1. Uma teoria epistemológica formal, por si só.

2. Uma lógica e sua semântica de suporte a uma teoria epistemológica.

Caso a compreendamos segundo a primeira perspectiva, teremos de especificar as circunstâncias em que a propriedade de onisciência lógica (expressa através de diversos tipos de fecho) representaria um problema para nosso esquema teórico. Para isso, teríamos que adicionalmente esclarecer as pretensões de explicação de nossa teoria, se seria uma teoria de perspectiva de primeira ou de terceira pessoa, suas pretensões de aplicação etc. Ou seja, a resposta à primeira pergunta estaria condicionada à resposta prévia de várias outras; daí porque não faria sentido perguntar acerca da validade ou invalidade de um determinado princípio de fecho epistêmico, antes que uma série de informações sobre a própria teoria responsável por esta análise fossem fornecidas. Essa é, ao meu ver, a mesma estratégia a ser adotada na epistemologia *mainstream*. A questão sobre a validade ou invalidade de princípios de fecho epistêmico só poderá ser posta após a especificação da lógica com a qual estamos trabalhando, acompanhada da informação acerca da sua perspectiva (se de primeira ou terceira pessoa), bem como de suas pretensões de aplicação – isto é, da especificação das situações que pretende modelar.

Como já foi visto, na segunda perspectiva, a lógica epistêmica não tem, por si só, a finalidade de interpretar as noções de conhecimento e crença; essa tarefa ficaria supostamente para a epistemologia. Deste modo, um modelo lógico-epistêmico até poderia interpretar noções como “conhecimento” e “crença”, mas sempre com base em uma teoria epistemológica que daria significado a essas noções. Nessa perspectiva, uma das finalidades de uma lógica epistêmica – entre várias outras, é claro – seria a de modelar (e ser coerente com) um modelo epistemológico qualquer; assim, uma lógica epistêmica que modela uma teoria epistemológica deve ser construída de modo que todos os seus resultados sejam compatíveis com os desta última. Isto é:

- Todos os teoremas da lógica epistêmica deveriam ser considerados proposições verdadeiras na teoria epistemológica;
- Quando traduzidas para a linguagem formal do modelo lógico-epistêmico, todas as proposições verdadeiras da teoria epistemológica seriam também verdadeiras (ou teoremas).

Com base em tudo o que foi mostrado até agora, fazemos a seguinte questão:

Na perspectiva da lógica epistêmica como modelo de suporte, a onisciência lógica seria um problema?

Nossa argumentação sugere a seguinte resposta: não necessariamente. A propriedade de onisciência surge como um problema dependendo da aplicação pretendida para lógica epistêmica de interesse. Considere-se, por exemplo, que há várias teorias epistemológicas; ou melhor, há vários modos de interpretar a noção de conhecimento e crença. Similarmente, há várias maneiras de se entender o ato de conhecer.

Cada teoria epistemológica é apenas um modelo que tenta descrever e compreender aquilo que observa. Sabemos que modelos são abstrações, idealizações, e que não necessitam, portanto, ser limitados a ponto de representar somente aquilo que pode ser apreendido pelos sentidos. Na verdade, idealizações são bastante comuns às ciências. É o caso, por exemplo, das linhas e pontos sem dimensões, das formas geométricas perfeitas da matemática, do vácuo absoluto da física ou dos átomos, entre outros. Deste modo, podemos considerar a propriedade da onisciência lógica também como uma idealização. Se uma teoria epistemológica postula, por qualquer razão que seja, a propriedade da onisciência lógica para seus agentes, uma lógica epistêmica tem o dever de assumir essa propriedade como um axioma de seu sistema. A aceitação do axioma $K_a(p \to q) \to (K_a p \to K_a q)$, juntamente com a regra da necessitação, nos permite demonstrar uma série de resultados interessantes – entre eles, a fórmula $(K_a p \wedge K_a q) \rightarrow K_a(p \wedge q)$ (E-CLOS 7). Mas como assim, interessante? Por qual motivo?

Para responder isso, primeiramente nos perguntamos se (E-CLOS 7), a saber, $(K_a p \wedge K_a q) \rightarrow K_a(p \wedge q)$, é um princípio aceitável. No primeiro capítulo, argumentamos que sim. Se um agente sabe que p é o caso e também sabe que q é o caso, então ele sabe que ambos são o caso. O que deve ser observado, contudo, é que a prova dessa fórmula é obtida em um sistema epistêmico padrão a partir axiomas e regras que representam propriedades de fecho. Fitting & Mendelsohn (1998, p. 69) oferecem a seguinte prova para essa fórmula[99]:

1. $p \rightarrow (q \rightarrow (p \wedge q))$	Tautologia
2. $K_a p \rightarrow K_a(q \rightarrow (p \wedge q))$	1, Regularidade
3. $K_a(q \rightarrow (p \wedge q)) \rightarrow (K_a q \rightarrow K_a(p \wedge q))$	Axioma $K_a(p \rightarrow q) \rightarrow (K_a p \rightarrow K_a q)$
4. $K_a p \rightarrow (K_a q \rightarrow K_a(p \wedge q))$	2,3 Lógica clássica
5. $(K_a p \wedge K_a q) \rightarrow K_a(p \wedge q)$	4, Lógica clássica

Apesar de ser uma fórmula amplamente aceita, já na sua prova foram utilizadas propriedades como a onisciência lógica. Vários outros casos similares a este ocorrem nos mais variados sistemas epistêmicos. A construção de idealizações, isto é, de sistemas epistêmicos com axiomas para agentes racionalmente perfeitos, nos permite esse tipo de discussão; podemos nos questionar porque fenômenos desta ordem ocorrem. Isso, ao que nos parece, ajuda cada vez mais a entender como funciona o pensamento logicamente limitado; isto é, comparando resultados "ideais" com aqueles que realmente obtemos de agentes "reais".

Além disso, vimos também, no capítulo anterior, que em algumas aplicações, o fecho sob implicação é considerado natural. A propriedade da onisciência passa a ser um problema quando esperamos que nosso modelo lógico satisfaça todas as capacidades e limitações de agentes reais – isto é, que simule todas as falhas de onisciência lógica. Vimos, entretanto, que isso está além do escopo da lógica. Problemas como o da onisciência lógica ocorrem quando os limites de aplicação de nossa teoria formal são extrapolados. Logo, a aplicação da teoria formal também

[99]A demonstração é feita originalmente em um sistema modal alético. Tendo isto, onde lê-se $\Box$, leia-se K_a. Na prova, também é utilizada uma regra derivada, chamada de "regra da regularidade". A prova dessa regra é obtida a partir de: 1. *Modus Ponens*; 2. axioma $K_a(p \rightarrow q) \rightarrow (K_a p \rightarrow K_a q)$; 3. regra da necessitação. A regra da regularidade tem a seguinte forma: $\dfrac{p \rightarrow q}{K_a p \rightarrow K_a q}$ $^{(Regularidade)}$.

determina seu sucesso ou insucesso em sua tentativa de modelagem. Para a lógica epistêmica, bem como para o problema da onisciência lógica, associamos pelo menos três perspectivas de investigação:

1. Simulação de motivos particulares de falha em onisciência lógica. Nesse caso, talvez até mais de um (recursos computacionais limitados e desconhecimento de regras relevantes, etc), mas não todos eles. Como vimos, às vezes falhamos em ser logicamente oniscientes por motivos que estão fora do alcance da lógica. Esta perspectiva sugere utilizar a lógica epistêmica para modelar tipos específicos de falha de onisciência lógica. É isso que viemos fazendo ao longo deste capítulo – e é isso o que os lógicos têm feito no que se refere ao problema da onisciência lógica (WHITSEY, 2003).

2. Utilização consciente da lógica epistêmica como uma abstração, uma idealização, para a obtenção de resultados e comparação desses resultados com aqueles aceitos para agentes reais. Nesse caso, vários problemas interessantes podem surgir. Entre eles, está aquele de investigar se podemos obter provas para teoremas amplamente aceitos sem contudo utilizar princípios questionáveis.

3. Utilização da lógica epistêmica como modelo lógico compatível com alguma teoria epistemológica. Neste caso, se nossa lógica epistêmica gerar resultados indesejáveis, devemos verificar se a tradução da teoria epistemológica para o modelo lógico foi efetuada corretamente. Se não, devemos rever nossa lógica epistêmica; se sim, devemos rever a teoria epistemológica. No entanto, essa linha de investigação tem sérias questões a responder. Uma delas, para começar, é justamente dizer se a formalização da "epistemologia *mainstream*" para a "epistemologia formal" é realmente possível – e, caso seja possível, se vale a pena fazê-lo. Como já mencionei, Hendricks (2006) vê nessa tentativa de aproximação grandes benefícios para a epistemologia contemporânea.

Concluirei esta seção com uma recomendação. Se alguém perguntar se a onisciência lógica é um problema, duas outras perguntas devem ser feitas a esse interlocutor, antes de qualquer resposta ser fornecida: "*Problema para quem?*", "*Em qual aplicação da lógica epistêmica você está pensando?*".

Felizmente, essas duas perguntas podem constituir motivos suficientes para que o referido sujeito desista da primeira... Seja lá o que for, já podemos perceber que a mera pergunta sobre a validade ou invalidade de um determinado princípio de fecho não faz sentido sozinha. Na verdade, até pode fazer, mas ela vem acompanhada de uma série de comprometimentos: tipo da lógica por trás da análise, perspectiva da desta lógica, suas pretensões de modelagem e aplicação etc.

2.6 E o fecho epistêmico, informalmente falando?

Como pudemos observar, os princípios de fecho são passíveis de investigação tanto na epistemologia formal quanto na informal. No primeiro capítulo, investigamos os desdobramentos da discussão acerca dos princípios de fecho a partir de uma perspectiva informal. Na maior parte deste capítulo, por outro lado, vimos os desdobramentos da discussão lógico-formal dos princípios de fecho – que ficaram evidenciados através da problemática da onisciência lógica. O que pudemos concluir disso? O seguinte: apesar das perspectivas dessas "duas" epistemologias serem diferentes, a estratégia para a investigação de princípios de fecho deve ser a mesma. O que isso quer dizer, então? Que, da mesma forma que não faz sentido perguntar sobre a validade ou invalidade de um esquema formal sem antes se explicitar a lógica que fará a análise, bem como as pretensões de modelagem da mesma, também não faz sentido perguntar sobre a validade de um princípio de fecho – epistemologicamente falando, é claro – sem que antes perguntas similares sejam respondidas. Se isso não for possível, entretanto, aquele que questiona, num plano epistemológico informal, a validade de um princípio de fecho, deve ao menos ter consciência dos comprometimentos lógico-epistemológicos que o próprio ato de colocar a questão acarreta.

Penso, a partir desses resultados, que a simplificação exagerada da análise dos princípios de fecho epistêmico é algo que deve ser evitado pela epistemologia informal se, com esta, quisermos aumentar nossa compreensão sobre tais princípios, bem como dos comprometimentos epistemológicos que a aceitação dos mesmos acarreta. Ora, por "simplificação exagerada" entendo justamente "a não-utilização de recursos (estratégias) da epistemologia formal (ou lógica epistêmica) na análise de princípios de fecho epistêmico".

Utilizemos, portanto, tal estratégia, e analisemos o fecho epistêmico com base na noção de "incognoscibilidade contingente". A noção de de "incognoscibilidade necessária" será utilizada para salientar as diferenças entre incognoscibilidade necessária e incognoscibilidade contingente, e mostrar que é justamente devido a essa diferença que hipóteses céticas, quando muito, podem ser consideradas apenas contingentemente incognoscíveis.

3 Incognoscibilidade necessária, contingente e análise dos princípios de fecho epistêmico

"Homo sum: humani nihil a me alienum puto."

(Terentius)

3.1 Introdução

Neste capítulo, os princípios de fecho discutidos no capítulo anterior, a saber, (E-CLOS 1), (E-CLOS 2), (E-CLOS 3) e (E-CLOS 6) serão analisados em sua relação com hipóteses céticas. Estas últimas, por sua vez, serão caracterizadas, através de definição, por "proposições contingentemente incognoscíveis", e serão diferenciadas de outras proposições (também caracterizadas através de definição) conhecidas como "necessariamente incognoscíveis"[1].

As proposições necessariamente incognoscíveis satisfazem aquilo que Rescher (2009) chama de incognoscibilidade necessária ou demonstrável, e foram estudadas com detalhe em trabalho de Fitch, publicado em 1963. Entretanto, esse tipo de incognoscibilidade difere daquela a ser trabalhada na seção 3.3, que chamo de "incognoscibilidade contingente". Será mostrado, por sua vez, que tanto as proposições *heavyweight* de Dretske quanto as hipóteses céticas satisfazem a noção

[1]A aplicação feita aqui do termo "proposição" pode ser, em certa medida, frouxa demais e geradora de confusões. Admitimos essa vagueza. Entretanto, esperamos do leitor que procure sempre ajuda no contexto em que o termo ocorre. Esperamos que as ocorrências desse termo sejam suficientemente clarificadas por aquilo que é alcançado através de sua aplicação. Entretanto, devemos lembrar que o operador de interesse epistemológico "conhece que..." é chamado de proposicional (ver nota 2, p. 15, cap. 1) porque remete a proposições, ou fatos descritos por estas proposições. Deste modo, ora podemos dizer "Zé Eduardo sabe que o Sol é uma estrela" – isto é, algo factual, pois ele conhece algo sobre o mundo – ou "Zé Eduardo sabe que P é o caso" (sendo P a proposição "o Sol é uma estrela").

de incognoscibilidade contingente e que, por isso, podem ser caracterizadas como "proposições contingentemente incognoscíveis".

Nas considerações sobre os resultados, chegamos à conclusão de que os princípios de fecho acima mencionados não são aplicáveis a situações em que atribuidores de conhecimento consideram hipóteses céticas como proposições contingentemente incognoscíveis, e sugerimos que tal estratégia da epistemologia formal, de análise de princípios de fecho a partir da perspectiva da modelagem e aplicação da lógica de interesse, seja adotada pela epistemologia informal.

3.2 O teorema de Fitch: incognoscibilidade necessária

Nós, seres humanos, somos conhecedores finitos; o conhecimento que temos sobre o mundo é limitado, e reconhecemos esta limitação. Por razões que talvez até desconheçamos, não somos ainda capazes de produzir respostas consensuais para várias questões científicas, religiosas, filosóficas etc. Somos, por assim dizer, epistemologicamente limitados. Apesar disso, não somos, de modo algum, inconformados com tal realidade; simplesmente aceitamos este fato em nossas vidas. Até podemos conjecturar sobre deuses, onisciência lógica etc., mas sabemos que tais conjecturas não se aplicam a nós. Ou seja, sabemos que não sabemos tudo. Proposicionalmente falando, se não sabemos de tudo, isso significa dizer que há ao menos uma proposição que não nos é conhecida. Até aí, tudo bem: existem proposições desconhecidas por nós. O que interessa neste capítulo, porém, é investigar as proposições incognoscíveis? Isto é, aquelas que, por definição, não podem ser conhecidas por ninguém. A pergunta que motiva a pesquisa sobre incognoscibilidade, portanto, é a seguinte: Existem proposições que estão fora do alcance cognitivo de qualquer ser humano?

Não é difícil compreender que a resposta para essa pergunta é um veemente "sim". Em 1963, Fitch (p.138-139) estabeleceu esta tese ao apresentar dois teoremas bastante curiosos:

Teorema 3.1. *Para cada agente que não é onisciente, existe uma proposição que esse agente não pode conhecer.*

Teorema 3.2. *Se existe alguma proposição que ninguém conhece (ou conheceu, ou irá conhecer) que é verdadeira, então existe uma proposição verdadeira que ninguém* *pode conhecer que é verdadeira.*

As demonstrações destes teoremas são bastante simples. Para isso, considere as seguintes proposições[2]:

P: A espada de César continha tungstênio.

Q: P é uma proposição verdadeira que o agente S não sabe que é verdadeira.

Agora, suponhamos um agente qualquer, S, tal que S não sabe que P seja verdadeira. Ora, se P é uma proposição desconhecida por S, então a proposição Q é incognoscível para S. Para isso, basta observar que Q é uma conjunção:

"*P é verdadeira*" e "*o agente S não sabe que P é verdadeira*".

Se, por acaso, aceitarmos que S conhece Q, então o princípio da distribuição (E-CLOS 8) nos dará "*S sabe que P é verdadeira*" e "*S sabe que ele (S) não sabe que P é verdadeira*"; mas daí, se tudo o que é conhecido deve ser verdadeiro (princípio da veracidade), então obtemos "*S não sabe que P é verdadeira*", o que é absurdo – já que havíamos aceito a verdade de "*S sabe que P é verdadeira*". Logo, existe uma proposição incognoscível para um agente particular arbitrário, a saber, S.

Observe ainda que esta demonstração vale para qualquer agente. Como S é um agente arbitrário, a demonstração pode ser generalizada, fornecendo-nos o seguinte resultado: para cada agente particular, existe pelo menos uma proposição que esse agente não pode conhecer. Para derivar este resultado, basta aceitarmos, através de uma hipótese nada exigente, que existe pelo menos uma proposição desconhecida para cada agente. Como – pelo menos no que concerne aos humanos – ninguém conhece todas as proposições (que são infinitas), o resultado de Fitch pode ser considerado um teorema. Ademais, seguindo o raciocínio de Rescher (2005, p. 18), imagine que t_1 é alguma proposição particular incognoscível para o agente S_1, t_2 para o agente S_2 e t_n para o agente S_n. Seja t^* a conjunção $(t_1 \wedge t_2 \wedge, ..., \wedge t_n)$. O resultado é que a conjunção t^* é uma proposição incognoscível para cada um dos agentes envolvidos. Logo, existem proposições que ninguém conhece e, além disso, que ninguém irá jamais conhecer[3].

[2]As demonstrações originais de Fitch são diferentes devido ao fato de ele ilustrar o operador de conhecimento como uma classe – isto é, a classe de proposições conhecidas. Para ilustrar a demonstração original, basta uma simples adaptação de linguagem formal

[3]O exemplo de Rescher, no entanto, é voltado para um tipo diferente de incognoscibilidade. Mas adiante, farei a distinção entre incognoscibilidade necessária e incognoscibilidade contingente. Ve-

É natural que questionemos a nós mesmos sobre o poder desses teoremas. Até onde podem ser aplicados? Para quem, efetivamente, a proposição t^* é incognoscível? Em seu livro sobre incognoscibilidade, Rescher (2009, p. 6) escreve:

> Agora, quando algum fato é tido como incognoscível, a questão que irá imediatamente surgir: para quem? Há várias perspectivas, especificamente[4]:
>
> - para um indivíduo;
> - para humanos em geral;
> - para seres inteligentes finitos, como um todo.

Rescher está entre aqueles que aceitam que os teoremas 3.1 e 3.2 de Fitch se aplicam, de fato, a qualquer ser inteligente que não é onisciente. Ele também observou algo bem interessante acerca dos fatos incognoscíveis: eles sempre podem ser apresentados como respostas para questões particulares. Por quê? Ora, suponha que você faça a seguinte pergunta a um agente:

Qual é um exemplo de um fato que você não conhece?

Certamente, o agente em questão não será capaz de fornecer uma resposta satisfatória para essa pergunta. Para fazê-lo, ele ou ela teria de conhecer o próprio fato, e isso é justamente o que está sendo proibido. Da mesma forma, se eu fizer a mesma pergunta a você, leitor, o resultado se repetirá; não será possível para você respondê-la. Apesar disso, observe que a questão é genuína. Ela ainda permanece. Alguém ainda poderia respondê-la em seu lugar! O fato de você não tê-la respondido não implica que ela não tenha uma resposta. Seguindo a mesma linha de raciocínio, encontramos a seguinte questão:

Qual é um exemplo de proposição que ninguém conhece?

---

remos que o referido exemplo é utilizado por Rescher para caracterizar esta última. Para isso, ele começa supondo proposições meramente "desconhecidas", ao invés de "necessariamente incognoscíveis" – como foi feito aqui. Entretanto, observei que a mesma estratégia pode ser aplicada às proposições necessariamente incognoscíveis; para isso, basta justamente utilizar proposições necessariamente incognoscíveis diferentes, particulares, para agentes também particulares. Quando perseguido de modo correto, tal raciocínio leva à proposição desejada: uma conjunção necessariamente incognoscível a todos os agentes, de modo geral.

[4] "*Now when some fact is said to be unknowable, the question will immediately arise: for whom? And there are various prospects here, specifically*:

- *for a given individual*;
- *for humans in general*;
- *for finite intelligent beings at large*".

Neste caso, ninguém pode fornecer uma resposta. O próprio ato de fornecer uma instância de uma proposição que ninguém conhece irá destruir o que está sendo pressuposto, a saber, que a proposição em questão é universalmente desconhecida. Apesar disso, não se pode negar a existência de tal proposição. Fatos incognoscíveis, bem como proposições incognoscíveis, existem; apenas não somos capazes de fornecer instâncias deles.

Mas isso pede um esclarecimento. Esses fatos e proposições incognoscíveis, bem como as questões irrespondíveis que acabamos de apresentar, têm uma forma bem particular: eles estão vinculados ao que Rescher chama de "incognoscibilidade necessária ou demonstrável" (2009, p.3) – o tipo de incognoscibilidade demonstrado por Fitch, com os teoremas 3.1 e 3.2.

Para os propósitos deste capítulo, devemos definir rigorosamente o termo "proposição necessariamente incognoscível". Esta definição, por sua vez, ser-nos-á bastante útil na análise e avaliação de princípios de fecho; além disso, também será vital para a tese de que as hipóteses céticas podem ser caracterizadas como "proposições contingentemente incognoscíveis". Ou seja, a noção de "proposição necessariamente incognoscível" é a contraparte necessária, e também complementa, a noção de "proposição contingentemente incognoscível". É justamente através desta estratégia – isto é, definir precisamente os dois tipos de incognoscibilidade – que será possível mostrar que as conhecidas hipóteses céticas – mencionadas e utilizadas no capítulo 1 – não podem ser caracterizadas como "proposições necessariamente incognoscíveis", mas tão somente como "proposições contingentemente incognoscíveis". Tal estratégia, que por si só já constitui importante avanço na discussão sobre ceticismo, resultará – em uma aplicação bem particular – na falha de alguns tipos de fecho. Chamarei esta aplicação (contexto) de "conjecturadores de hipóteses céticas"[5].

Abaixo, encontra-se a incognoscibilidade (proposicional) necessária.

[5] O ato de pensar a problemática do fecho a partir da perspectiva "aplicação-contexto" é inspirado, como se pode observar, na epistemologia formal; ou seja, utilizei exatamente a mesma estratégia do capítulo 2, quando então tratei da problemática da onisciência lógica. Utilizando a mesma estratégia dos lógicos, sugiro neste trabalho a análise de princípios de fecho a partir de uma perspectiva bem particular, aquela que limite tanto o fecho quanto o conceito de validade ao contexto de trabalho – contexto este que deve, desde já, ser previamente definido, juntamente com as intenções de modelagem e de utilização do referido contexto. Entretanto, a discussão do contextualismo epistêmico é bastante ampla e controversa, de modo que a mera pressuposição de uma perspectiva contextualista (isto é, sem suma consideração pormenorizada de seus problemas) possa parecer algo irresponsável. Entretanto, para evitar mais complicações, pensemos aqui em "contexto" como sendo uma "situação aplicável de um princípio lógico" ou, para simplificar, uma "aplicação".

Definição 3.1. ***Incognoscibilidade necessária.*** *Seja P uma proposição qualquer e S um agente qualquer. Dizemos que P é uma proposição necessariamente incognoscível para S se, e somente se, P é uma proposição composta do tipo $K_S(Q \wedge \neg K_S Q)$, na qual Q é uma proposição que o agente S desconhece*[6].

Como se pode notar, P é uma proposição incognoscível por razões puramente lógicas. Já foi mencionado anteriormente que Rescher caracteriza esse tipo de incognoscibilidade como "necessária" ou "demonstrável"; aquele tipo de incognoscibilidade em que o significado de P é totalmente irrelevante, já que P é incognoscível "em princípio", isto é, com base simplesmente em princípios lógicos. Não obstante de ser ou não uma hipótese cética, P é aquele tipo de proposição que está e permanecerá fora do alcance cognitivo de qualquer ser inteligente que seja racionalmente limitado – isto é, que não seja onisciente.

As proposições necessariamente incognoscíveis, por sua vez, diferem significativamente das proposições contingentemente incognoscíveis. Argumentar-se-á, aqui, que estas últimas constituem, entre outras, aquelas proposições que conhecemos por "hipóteses céticas"[7].

3.3 Incognoscibilidade contingente

3.3.1 As perguntas contingentemente irrespondíveis

É bem provável que nenhum ser humano, neste exato momento, esteja apto a fornecer respostas conclusivas a quaisquer das perguntas abaixo:

(1) De quanta água Pôncio Pilatos precisou para lavar suas mãos antes de dar a condenação de Jesus?

[6]A proposição $K_S(Q \wedge \neg K_S Q)$ é lida do seguinte modo: "S sabe que a seguinte proposição é verdadeira: Q é verdadeira mas ele, S, não sabe que Q é verdadeira".

[7]Além das hipóteses céticas, as proposições *heavyweight* também podem ganhar a mesma caracterização, a saber, de proposições contingentemente incognoscíveis. Em concordância com McBride (2009), essa ideia também ajuda a corroborar a tese de que o chamado "desafio de Dretske" encontra-se, até o presente momento, em aberto. Entretanto, este é não um comprometimento obrigatório. Nem as proposições céticas nem as proposições *heavyweight* precisam ser consideradas proposições contingentemente incognoscíveis; porém, num contexto filosófico, em que os padrões para a aquisição e manutenção do conhecimento são muito altos, a possibilidade de tal caracterização não deixa de ser útil. A noção de incognoscibilidade contingente é apropriada para ser aplicada a situações em que tais padrões são muito exigentes. Em outras palavras, a noção de incognoscibilidade contingente tem grande utilidade na determinação do *status* epistêmico de uma proposição em situações em que os padrões para a aquisição e manutenção do conhecimento são altos.

(2) Qual a quantidade exata de cicuta que Sócrates ingeriu antes de morrer?
(3) Quantos seres humanos estavam vivos na época da crucificação de Jesus?
(4) Qual é o número exato de estrelas no universo?
(5) Nosso universo está contido dentro de uma casca de noz?

Muitos não oferecerão resistência à tese de que essas perguntas são, por assim dizer, irrespondíveis – pelo menos, boa partes delas. De fato. Suponha, por exemplo, que escavações recentes na cidade de Jerusalém nos tenham revelado a existência de um novo documento que descreve, com detalhes, todos os aspectos do julgamento de Jesus. Assim, entre as informações contidas, está a de que Pôncio Pilatos utilizou o equivalente a 500ml de água para lavar suas mãos antes de condenar Jesus à morte. Ora, mesmo que tal documento existisse, ainda assim teríamos a controversa questão de saber se o que ele conteria seria fidedigno. Além disso, considerando que, hoje em dia, existem métodos mais precisos para medir as quantidades de água em potes, e que tais métodos não nos estão disponíveis para tratar da questão de Pôncio Pilatos, chegamos à conclusão de que há somente três alternativas para lidar com tal problema: (1) acreditar no testemunho do documento, (2) desconsiderar o testemunho do documento ou (3) suspender temporariamente o juízo, enquanto se procura por mais evidências. Entretanto, observe que a opção (3) não parece muito promissora; mesmo que haja mais evidências, elas sempre compartilharão da mesma característica da primeira evidência: a de que o julgamento de Jesus se passou há muito tempo atrás, e de que as supostas "evidências" são muito antigas e, além disso, passíveis de inúmeras interpretações – algumas até mesmo excludentes. Ou seja, tal questão se perde num emaranhado de inúmeras outras, igualmente relevantes para a determinação de uma resposta conclusiva para a primeira. Assim, é natural que, quando deparados com questões como (1), simplesmente aceitemos que sua resposta encontra-se distante demais de nós, talvez até de modo inalcançável – coisa que, igualmente, não podemos afirmar de modo conclusivo.

Note-se que as demais questões – isto é, (2)-(5) – comportam-se de modo similar. Todavia, observe-se também que não podemos negar que suas respectivas respostas existam. Pôncio Pilatos precisou de uma quantia finita de água para lavar suas mãos antes de dar a Jesus a sentença de morte – quantia que poderia ser determinada naquele momento histórico, e que poderia ser expressa através de uma proposição. Similarmente, Sócrates bebeu uma quantia finita e definida de cicuta antes de morrer. Além disso, até onde sabemos, existia uma quantia finita de seres humanos vivos, tanto na época de Jesus quanto na de Sócrates. O número de estrelas no universo é finito ou infinito. Nós não sabemos a resposta. Supondo que ele fosse finito, haveria um número determinado que poderia ser utilizado para responder corretamente a pergunta “Qual é o número exato de estrelas no universo?”. Sobre esta última questão, diríamos apenas que não fomos espertos o suficiente para descobrir a resposta!

A essa altura, já podemos perceber que certas questões, apesar do fato de possuírem suas respectivas respostas, são irrespondíveis para nós no momento exato em que as colocamos – isto é, no momento exato em que as estamos pensando. A elas darei o nome de “perguntas contingentemente irrespondíveis”; já suas respostas, por sua vez, recebem o nome de “proposições contingentemente incognoscíveis”. Mas quando é que uma pergunta deve ser considerada contingentemente irrespondível? Primeiramente, uma pergunta pode ser irrespondível para um agente qualquer S, por exemplo, simplesmente porque:

- O agente S não é esperto o suficiente para descobrir a resposta.
- O agente S possui, digamos, um conjunto incompleto de regras de inferência que o auxiliariam a derivar a referida resposta.
- As evidências de que S precisa para derivar a resposta requerida não estão disponíveis[8].

Com base nisso, fica fácil perceber porque uma determinada pergunta pode ser considerada “contingentemente irrespondível” para um agente qualquer, S. As questões (1)-(5) são contingentemente irrespondíveis para um agente S exatamente quando as razões que as tornam irrespondíveis são contingentes; mas não apenas isso: além de contingentes, a indisponibilidade de evidências necessárias para sua resposta deve ser garantida, isto é, consensualmente aceita pelos agentes que compartilham tais questões, num contexto conversacional determinado. Em outras palavras, dentro de um contexto determinado, os agentes que colocam questões como essas uns aos outros devem estar em consenso sobre a indisponibilidade de evidências necessárias para a resolução desses problemas, bem como da baixa probabilidade de adquiri-las. Na tentativa de captar essa ideia, a seguinte definição para o termo “pergunta contingentemente irrespondível” pode ser proposta:

[8]Rescher (2009, p. 2) aplica o mesmo raciocínio a fatos. Na verdade, observa-se que a mesma explicação funciona para as questões irrespondíveis.

Definição: ***pergunta contingentemente irrespondível.*** *Seja P uma pergunta qualquer, S um agente qualquer e E um conjunto não-vazio de evidências (proposições) cujo conhecimento é necessário para uma resposta correta à pergunta P. Dizemos que P é contingentemente irrespondível para S, num momento qualquer t (em que a referida questão é colocada), se, e somente se:*

1. *Existe pelo menos uma proposição Q do conjunto E que é consensualmente considerada, no contexto conversacional em uso, uma proposição contingentemente incognoscível para S, no momento t. Ou*
2. *a resposta para P é uma proposição contingentemente incognoscível.*

Agora, à luz desta definição, voltemos à questão (1). Segundo nossa definição, (1) é contingentemente irrespondível para nós, neste exato momento, porque todos concordamos que não existe qualquer informação disponível acerca da quantidade exata de água que Pôncio Pilatos utilizou para lavar as mãos antes de condenar Jesus; não existe qualquer relato preciso sobre o fato, qualquer texto com credibilidade histórica que tenha chegado até nós e que afirme, com todos os detalhes necessários, qual foi a quantidade exata de água utilizada por Pilatos. Ou seja, concordamos que, neste exato momento, existe uma proposição que responde à pergunta (1) e que, para nós, é contingentemente incognoscível.

A definição acima explicaria o por quê de (1) ser considerada contingentemente irrespondível, não fosse por um detalhe: cometer a falácia de circularidade – na medida em que utiliza, em sua ação definidora, o termo "proposição contingentemente incognoscível", um conceito evidentemente muito próximo. De fato, ao observar melhor a questão, não será difícil perceber que "questões contingentemente irrespondíveis" e "proposições contingentemente incognoscíveis" são conceitos que complementam um ao outro. Apesar disso, é perfeitamente possível oferecer definições de ambos sem que incorramos em petição de princípio. Para isso, basta definir, primeiro, o conceito de "proposição contingentemente cognoscível", e fazer do termo "pergunta contingentemente incognoscível" um conceito derivado. Sigamos, portanto, essa estratégia.

3.3.2 As proposições contingentemente incognoscíveis

Grosso modo, uma proposição contingentemente incognoscível é aquela que é incognoscível por razões contingentes. Mas o que isso quer dizer? Para entendêlo, não é preciso ir muito longe. Imagine novamente, por exemplo, o caso de Pôncio Pilatos. No dia do julgamento de Jesus, Pilatos utilizou uma quantidade finita de água para lavar as mãos. Ora, essa quantidade finita de água, que obviamente desconhecemos, pode ser determinada por uma proposição qualquer, *P*, do tipo *Pôncio Pilatos utilizou "X" quantidade de água para lavar as mãos, no dia do julgamento de Jesus*. Apesar de não conhecermos a proposição que estamos procurando, isto é, aquela que corretamente determina do valor de "*X*", não podemos discordar de que tal proposição exista. Tal proposição nos é, pura e simplesmente, inacessível no momento.

Porém, e se as coisas tivessem sido diferentes? Ora, se um conjunto especial de condições pudesse ser satisfeito, talvez a proposição em questão pudesse ser conhecida. Alguém poderia ter observado o evento do julgamento de Jesus e o documentado de alguma forma. Melhor ainda, e se a máquina do tempo fosse inventada? Ora, se essa condição fosse satisfeita, poderíamos voltar no tempo e conhecer a proposição em questão por nós mesmos. Quem sabe? Não existe nada de errado com o seguinte contrafactual:

(C-FACT 1): "Se as coisas tivessem sido diferentes, P poderia ser conhecida por nós".

Ou seja, se determinadas condições – necessárias para o conhecimento da proposição *P* – pudessem ser satisfeitas, *P* seria uma proposição conhecida por nós. Naturalmente, as objeções contra este contrafactual seriam concentradas nestas "coisas diferentes"; isto é, argumentar-se-ia sobre a plausibilidade das condições a serem satisfeitas. No entanto, isso não é relevante. O fato de simplesmente permitir a possível satisfação – seja ela provável ou não – deste contrafactual torna a proposição em questão contingentemente incognoscível. Em outras palavras, a proposição em questão é incognoscível, mas por razões contingentes; isto é, na medida em que (C-FACT 1) não for satisfeito, *P* permanece (contingentemente) incognoscível. Em contrapartida, neste caso específico, se (C-FACT 1) fosse satisfeito, a proposição *P* deixaria de ser contingentemente incognoscível e tornar-se-ia, portanto, conhecida por nós. É nisto em que consiste uma proposição contingentemente incognoscível.

Certamente, várias objeções poderiam ser feitas. Uma delas é justamente questionar a própria utilização do termo "incognoscível". Afinal de contas, se existe uma possibilidade, mesmo que remota, de proposições como essas serem conhecidas por nós, que sentido faz chamá-las de incognoscíveis? "Incognoscível" não seria aquilo que, por definição, não pode ser conhecido?

Na verdade, a resposta a esta questão irá depender justamente da interpretação específica que se tem do referido termo. Esta objeção deixa algo claro: que o termo "incognoscível" está sendo interpretado em seu sentido forte. Nesta perspectiva, "incognoscível" é aquilo que não pode ser conhecido de modo algum. Tal concepção de incognoscibilidade vale, de fato, para muitos contextos. No entanto, em seu sentido forte, o referido termo não permite diferenciar entre proposições logicamente incognoscíveis – como é o caso das proposições estudadas por Fitch – e proposições como "César comeu 360 gramas de queijo em seu último café da manhã" ou "não sou um cérebro numa cuba". Esta última, como sabemos, é considerada "incognoscível" pelos céticos – e por epistemólogos como Dretske, por exemplo[9]. Isto é, mesmo concordando acerca de sua incognoscibilidade, as respectivas causas de sua incognoscibilidade diferem amplamente. É justamente isso que a estratégia da distinção entre incognoscibilidade necessária e incognoscibilidade contingente procura captar. Além disso, ela considera, também, a relevância do contexto na determinação da (in) cognoscibilidade (contingente) de proposições. Por exemplo, existe uma proposição que descreve a quantidade exata de carboidratos ingeridos por Júlio César em seu derradeiro café da manhã. Esta proposição, como sabemos, é-nos desconhecida. Além disso – e muitos hão de concordar – tal proposição nunca nos será conhecida, assim como muitos concordam que também nunca nos será conhecida a proposição que identifica, corretamente, o assassino conhecido como "Jack, o estripador".

Em outras palavras, em nosso atual contexto conversacional, nós – agentes epistêmicos em questão – estamos em consenso acerca da falta de evidência – bem como da baixa probabilidade em adquiri-las – para as proposições que estamos considerando, a saber, aquela que determina com sucesso a quantidade exata de carboidratos ingeridos por César, bem como aquela que determina com sucesso o nome verdadeiro de "Jack, o estripador". Tais proposições são aquilo que podemos denominar por "contingentemente incognoscíveis".

Entretanto, para evitar futuras complicações, devemos oferecer uma definição menos vaga para a incognoscibilidade contingente. Para isso, comecemos primeiro com a seguinte definição:

[9]Na próxima seção, a necessidade de distinguir "incognoscibilidade necessária" e "incognoscibilidade contingente" tornar-se-á mais clara. É justamente ela que irá impedir-nos de confundir hipóteses céticas com proposições logicamente incognoscíveis.

Definição 3.2. *Contrafactual da contingência.* *Seja P uma proposição qualquer e S um agente qualquer, que não conhece a proposição P; seja (C-FACT) um contrafactual qualquer e C um conjunto não-vazio de condições de (C-FACT). Dizemos que (C-FACT) é um contrafactual da contingência para P, relativamente a S, se, e somente se, possui a seguinte forma: Se o conjunto C de condições fosse satisfeito, P seria uma proposição conhecida por S.*

Como vimos, os "contrafactuais da contingência" são extremamente importantes na identificação de proposições contingentemente incognoscíveis. De posse desse importante conceito, continuemos com a definição de incognoscibilidade (proposicional) contingente:

Definição 3.3. *Proposição contingentemente incognoscível.* *Seja P uma proposição qualquer, S um agente qualquer e (C-FACT) um contrafactual da contingência, nos moldes da definição 3.2. Dizemos que P é contingentemente incognoscível para S, num momento qualquer t (em que a referida questão é colocada), se, e somente se:*

1. *S não conhece P em t;*
2. *P é consensualmente considerada, no contexto conversacional em uso – isto é, em que S está inserido – uma proposição que carece de evidências de suporte e, além disso, a probabilidade de se encontrar tais evidências, no momento t, é considerada muito baixa*[10]*;*
3. *(C-FACT) é um contrafactual da contingência para P.*

Através desta definição, podemos explicar a incognoscibilidade contingente de várias proposições. Para isso, consideremos novamente o famoso caso do julgamento de Jesus. Suponhamos que, naquela ocasião, Pôncio Pilatos precisou de cerca de 500ml água para lavar as mãos antes de condenar Jesus à morte. Deste modo, uma proposição que corretamente asserta este fato poderia ser:

P_{ci1}: *Pôncio Pilatos utilizou 500ml de água para lavar as mãos, antes de condenar Jesus à morte.*

[10] Ou seja, consensualmente, considera-se que as evidências necessárias para o conhecimento da proposição *P* não estão disponíveis; que também não há como adquiri-las no momento em questão e, além disso, que a probabilidade de que tais evidências sejam adquiridas em qualquer tempo futuro é (consensualmente) considerada muito baixa.

Agora, considere um desavisado agente, S, que não conhece esta proposição. Ou seja, S encontra-se em completa ignorância acerca do fato que a referida proposição asserta – isto é, que Pôncio Pilatos utilizou "500ml" de água para lavar as mãos, antes de condenar Jesus à morte. Nós, que nesta ocasião somos atribuidores de valor de verdade à proposição *"S não conhece P_{ci1}"*, em nossa posição privilegiada, constatamos que:

1. S não conhece P_{ci1}.

Como se pode observar, uma condição da definição 3.3 já foi satisfeita, a saber, aquela da ignorância do agente S acerca da proposição de nosso interesse – isto é, P_{ci1}. Agora, precisamos explicitar o porquê de estarmos em uma posição epistemicamente privilegiada em relação a S.

Ora, as razões para isso podem ser muitas. Pode-se supor, por exemplo, que tivemos acesso a uma fonte de conhecimento que julgamos confiável e que ela atesta que Pilatos utilizou 500ml de água antes de condenar Jesus. Apesar da improbabilidade, poderia também ser o caso de sermos *highlanders*, isto é, seres imortais, e que estivemos presentes ao julgamento de Jesus. Além disso, poderíamos ainda ser possuidores de uma máquina do tempo, e que a utilizamos para visitar o julgamento de Jesus etc. Resumindo, existem inúmeras circunstâncias logicamente possíveis que permitem explicar, neste caso específico acerca do julgamento de Jesus, nossa posição epistemicamente privilegiada em relação ao agente S.

Similarmente, além de ser possível conjecturar várias formas de privilégio epistêmico que mantemos em relação ao agente S, também é possível conjecturar acerca da improbabilidade de S "ascender" a esta posição epistemicamente privilegiada. Para o caso de sermos *highlanders*, basta supor (i) que S é um simples mortal, que (ii) ignora plenamente a existência de *highlanders* e (iii) que nós, enquanto *highlanders*, guardamos em segredo (e com a própria vida) nossas identidades. Para o caso da máquina do tempo, podemos pensar em algo similar: isto é, de que guardamos em segredo e com a própria vida, caso necessário, sua existência etc. Já o caso da "fonte confiável", apesar de mais próximo de nosso contexto atual, pode comportar-se de modo similar. Basta imaginar que a referida fonte é uma famosa revista científica especializada que o agente S desconhece plenamente, por não ter acesso a esse tipo de literatura nem a preparação especializada necessária para apreciá-la; ou que, por suas convicções filosóficas, religiosas, políticas etc., mesmo que tivesse acesso a uma fonte dessas, não a aceitaria como confiável. Em outras palavras, constatamos também que:

2. P_{ci1} é consensualmente considerada, no atual contexto conversacional – isto é, em que S está inserido – uma proposição que carece de evidências de suporte e, além disso, a probabilidade de se encontrar tais evidências, no momento t, é considerada muito baixa.

Agora, para finalizar, observe-se também que existem contrafactuais da contingência para cada um dos casos explicitados no parágrafo anterior. Isto é:

(C-FACT 1) Se um *highlander*, por razões pessoais, resolvesse revelar seu segredo a S, S saberia que Pilatos utilizou 500ml de água antes de condenar Jesus à morte.

(C-FACT 2) Se um determinado inimigo meu, querendo causar-me aborrecimentos, roubasse minha máquina do tempo e levasse S com ele para assistir ao julgamento de Jesus, S saberia que Pilatos utilizou 500ml de água antes de condenar Jesus à morte.

(C-FACT 3) Se S resolvesse ser um cientista e o conseguisse, e tivesse acesso a revistas técnico-científicas acerca de acontecimentos históricos, ele saberia que Pilatos utilizou 500ml de água antes de condenar Jesus à morte.

Não é difícil notar que (C-FACT 1), (C-FACT 2) e (C-FACT 3) são todos contrafactuais da contingência para a proposição P_{ci1}. Assim, no atual contexto conversacional, também constatamos que:

(C-FACT 1-3) são contrafactuais da contingência para P_{ci1}.

Logo, dada a satisfação das três condições exigidas na definição 3.3, conclui-se que a proposição P_{ci1} é contingentemente incognoscível. É fácil perceber que proposições similares a P_{ci1} satisfazem a definição 3.3. Como vimos na seção anterior (sobre incognoscibilidade necessária), é logicamente impossível explicitarmos, para nós mesmos, quais são as proposições que nos são contingentemente incognoscíveis. O próprio ato de as explicitar já exclui o fato delas serem incognoscíveis para nós. Isso ocorre porque, neste caso, o acesso epistêmico privilegiado – que é necessário neste caso – nos é automaticamente negado com relação às respectivas proposições. Apesar disso, a definição 3.3 é uma maneira bastante prática para explicar porque determinadas proposições são, de forma contingente, inacessíveis epistemicamente a certos agentes. Resta-nos saber, no entanto, se podemos utilizar a definição 3.3 tanto para as proposições *heavyweight* de Dretske quanto para famosas hipóteses céticas como, por exemplo, dos cérebros em cubas e marionetes do gênio maligno cartesiano. Todavia, antes de executar essa tarefa, precisamos distinguir bem entre incognoscibilidade necessária e incognoscibilidade contingente; isto é, precisamos mostrar porque proposições necessariamente incognoscíveis são tão diferentes daquelas consideradas contingentemente incognoscíveis.

É justamente essa separação entre esses dois tipos de incognoscibilidade que nos permitirá demonstrar, do ponto de vista epistemológico, a caracterização tanto das proposições *heavyweight* quanto das hipóteses céticas como proposições contingentemente incognoscíveis. Ou seja, após a correta separação entre as noções de "incognoscibilidade necessária" e "incognoscibilidade contingente", ser-nos-á possível mostrar que as hipóteses céticas, bem como as proposições *heavyweight*, estão sob a ação definidora da definição 3.3.

3.4 Incognoscibilidades necessária e contingente

A essa altura, a diferença entre incognoscibilidade necessária e incognoscibilidade contingente talvez já tenha ficado clara por si só. Certamente, um bom começo é observar que, em filosofia, não é à toa que os filósofos têm o cuidado de utilizar termos diferentes para designar coisas diferentes. Ademais, a ambiguidade de termos em filosofia é sempre um preço muito alto a ser pago. Porém, para afugentar qualquer sombra de dúvida acerca dessas duas noções, utilizarei esta seção para diferenciar, de modo mais preciso, as duas formas de incognoscibilidade já mencionadas. Para isso, observe-se novamente a definição de proposição necessariamente incognoscível:

Seja P uma proposição qualquer e S um agente qualquer. Dizemos que P é uma proposição necessariamente incognoscível para S se, e somente se, P é uma proposição composta do tipo $K_S(Q \wedge \neg K_S Q)$, na qual Q é uma proposição que o agente S não conhece.

Como já vimos anteriormente, há algo bastante saliente nesta definição: a impossibilidade lógica de P ser conhecida pelo agente S. Para provar essa impossibilidade, basta raciocinar por absurdo – como já foi feito. Suponha que S conheça P. Ora, sabendo-se que P é a conjunção de duas outras proposições, a saber, Q e $\neg K_S Q$, por (E-CLOS 8) temos que S conhece tanto Q quanto $\neg K_S Q$. Ora, se S sabe

$\neg K_S Q$, então ele sabe que desconhece Q – o que é absurdo, já que S conhece Q. Logo, é falso afirmar que S conhece $K_S(Q \wedge \neg K_S Q)$. E, assim, temos que $K_S(Q \wedge \neg K_S Q)$ é demonstrativamente (e, portanto, logicamente) incognoscível. Isso é o caso justamente porque a hipótese de sua cognoscibilidade implica em absurdo.

Repare, também, que o que torna P necessariamente incognoscível é o desconhecimento da proposição Q. A falta de conhecimento de Q é requerida pela própria definição. Essa requisição, como se pode observar, não é negociável. Sem ela, a definição 3.1 nem mesmo faria sentido. Para que haja incognoscibilidade necessária para um agente qualquer, é preciso antes que esse agente ignore alguma proposição. No entanto, a proposição que S ignora não precisa, por sua vez, ser uma proposição necessariamente incognoscível. Na verdade, ela nem mesmo precisa ser contingentemente incognoscível. Ela pode, pura e simplesmente, ser uma proposição qualquer que o agente desconhece.

Considere, por exemplo, a proposição "*O Fiesta é um carro fabricado pela Ford.*" Como podemos constatar, esta proposição é conhecida pela maioria dos brasileiros. Porém, mesmo para quem não a conhece, ela pode facilmente ser conhecida: basta que se pergunte a um funcionário da *Ford* ou simplesmente a um "*car guy*" (fanático por carros). Entretanto, o jogo muda para um agente que não conhece esta proposição. Se por acaso existir no Brasil um agente que desconheça que o *Fiesta* é um carro fabricado pela *Ford*, então a proposição abaixo será necessariamente incognoscível para este agente:

O Fiesta é um carro fabricado pela Ford, mas eu não sei disso.

Como se pode observar, ao se aceitar que o referido agente conheça a proposição em questão, somos obrigados também a aceitar que ele não a conhece. Logo, é absurdo e portanto logicamente impossível afirmar que o referido agente possa conhecer a proposição em questão. Para que isso pudesse ocorrer, teríamos, primeiro, de aceitar que este agente conhecesse a proposição "*O Fiesta é um carro fabricado pela Ford*". Porém, é justamente isso que a definição 3.1 requer que o agente não saiba. Logo, o desconhecimento de uma proposição como Q, por exemplo, é algo logicamente exigido – sob pena de se cometer um absurdo – pela noção de incognoscibilidade necessária. Assim, como já afirmamos, uma proposição pode ser considerada logicamente incognoscível quando ela é incognoscível por razões puramente lógicas – sob pena de absurdo em caso contrário. Fora isso, a incognoscibilidade da proposição P na definição 3.1 não depende de qualquer outra coisa: e é justamente por isso que é incoerente afirmar que ela seja contingente. Ou seja, é simplesmente incorreto confundir o tipo de incognoscibilidade da definição 3.1 com aquele encontrado na definição 3.3; razão pela qual devemos utilizar termos diferentes para esses dois tipos de incognoscibilidade. Para que sejamos mais claros, analisemos agora a definição 3.3:

Seja P uma proposição qualquer, S um agente qualquer e (C-FACT) um contrafactual qualquer. Dizemos que P é contingentemente incognoscível para S, num momento qualquer t (em que a referida questão é colocada), se, e somente se:

1. *S não conhece P;*

2. *P é consensualmente considerada, no contexto conversacional em uso – isto é, em que S está inserido – uma proposição que carece de evidências de suporte e, além disso, a probabilidade de se encontrar tais evidências, no momento t, é considerada muito baixa*[11];

3. *(C-FACT) é um contrafactual da contingência para P.*

Podemos notar que esta definição depende, por sua vez, de outra: a definição 3.2, de "contrafactual da contingência." Ora, é exatamente este tipo de contrafactual que caracteriza as proposições contingentemente incognoscíveis. Na definição 3.3, é justamente o cláusula 3, do contrafactual da contingência, que garante a possibilidade lógica da proposição P ser conhecida – possibilidade esta proibida, logicamente, pela noção de incognoscibilidade necessária. E o que é mais interessante: a cláusulas 1 e 3 não são, de modo algum, incoerentes. Na verdade, o fato de elas não serem incoerentes é exatamente o que torna a proposição P desta definição contingentemente, e não necessariamente, incognoscível. Isso porque, como pode ser observado, a possibilidade lógica de satisfação do condicional da contingência permanece sempre aberta. Em outras palavras, por razões contingentes – referentes ao nosso mundo físico etc. – a proposição P não pode ser conhecida; porém, por razões lógicas, a cognoscibilidade de P não é proibida. Logo, a incognoscibilidade de P se dá por razões meramente contingentes; o que a faz, portanto, uma proposição contingentemente incognoscível.

[11]Ou seja, é consensual que as evidências necessárias para o conhecimento da proposição P não estão disponíveis; que também não há como adquiri-las no momento em questão e, além disso, que a probabilidade de que tais evidências sejam adquiridas em qualquer tempo futuro é (consensualmente) considerada muito baixa.

Acredito que, a essa altura, os dois tipos de incognoscibilidade estejam claramente diferenciados. À primeira vista, talvez, a estratégia de diferenciar essas duas formas de incognoscibilidade pode parecer desnecessária. Afinal de contas, nos mais variados contextos, os interlocutores parecem fixar por si mesmos a interpretação mais apropriada à circunstância.

Mas isso nem sempre ocorre. Na verdade, muitos concordariam que, na maioria das vezes – e principalmente em discussões filosóficas – grandes disputas ocorrem justamente porque não há acordo em relação ao significado dos termos empregados. Um bom exemplo disso é o problema do fecho epistêmico. Como já foi colocado no capítulo anterior, muitos dos problemas relativos ao fecho epistêmico – pelo menos no que se refere à questão da validade – poderiam ser evitados caso fosse aplicado ao conceito de validade a mesma estratégia já utilizada na epistemologia formal ou lógica epistêmica – a definição clara do conceito de validade e sua restrição ao contexto da lógica de interesse. Ou seja, a clara definição dos conceitos utilizados podem evitar, e muito, discussões longas e cansativas sobre a validade ou invalidade de certos princípios lógicos.

Apesar disso, a distinção das "incognoscibilidades" não está apenas condicionada ao problema do fecho epistêmico que, à luz desta distinção, será analisado nas próximas seções. A estratégia de diferenciar as noções de incognoscibilidade necessária e incognoscibilidade contingente, neste momento, tem um objetivo mais imediato: o da caracterização das proposições *heavyweight* e das hipóteses céticas clássicas da epistemologia informal como proposições contingentemente incognoscíveis. De fato, acredito que, utilizando a definição 3.3 – de proposição contingentemente incognoscível – é possível demonstrar que tanto as proposições *heavyweight* de Dretske, quanto hipóteses céticas como "não sou um cérebro numa cuba" podem ser caracterizadas como proposições contingentemente incognoscíveis.

Todavia, ressalta-se aqui que tal caracterização não precisa ser imposta a essas classes de proposições. Conhecemos suficientemente a discussão sobre proposições *heavyweight* e sobre o ceticismo para saber que a questão da cognoscibilidade ou incognoscibilidade de hipóteses céticas e proposições *heavyweight* está longe de chegar ao fim. Reconheço isso. Entretanto, ofereço àqueles que já sustentam (ou pelo menos aceitam) a incognoscibilidade de hipóteses céticas – como o próprio Dretske, por exemplo – uma nova forma de compreender tais proposições. Uma maneira de explicar qual é o tipo, a causa e modo pelo qual tais proposições são consideradas "incognoscíveis". Em outras palavras, a distinção entre incognoscibilidade necessária e contingente pode ser de grande valia para aquele que já aceita que nunca poderemos saber ao certo se somos ou não cérebros em cubas, marionetes do gênio maligno ou se o mundo externo existe. A partir daí, a invalidade de alguns princípios de fecho em certos contextos epistemológicos – isto é, de conjecturadores de hipóteses céticas – será, ao meu ver, uma posição defensável.

Assim, dando prosseguimento à discussão, investiguemos agora a relação existente entre as noções de incognoscibilidade contingente e proposição *heavyweight*.

3.5 *Heavyweightness* e incognoscibilidade contingente

3.5.1 Definindo fatos, proposições e questões *heavyweight*

Como vimos no primeiro capítulo, uma proposição *heavyweight* é definida como aquela que não é cognoscível pela percepção, mesmo que esta última seja auxiliada pela razão. Dito isto, observa-se que todas as proposições abaixo satisfazem esta definição:

H_1: Não sou um cérebro numa cuba.

H_2: Não sou uma marionete do gênio maligno cartesiano.

H_3: O mundo externo existe.

H_4: Aquela zebra não é uma mula disfarçada.

H_5: Não sou um prisioneiro da *Matrix*.

Suponha – nem que seja por um breve momento – que todas as proposições supracitadas sejam verdadeiras. Agora, observe que, se há proposições *heavyweight* verdadeiras (tome-se H_1-H_5 como exemplos), então é natural pensar que também há "fatos" e "questões *heavyweight*". Estes últimos podem ser definidos do seguinte modo:

Definição 3.4. ***Fato heavyweight***: *Um fato heavyweight é uma verdade que não é cognoscível pela percepção, razão, testemunho, introspecção ou memória*[12].

[12]Deste modo, um fato heavyweight não é cognoscível por qualquer fonte convencional de conhe-

Definição 3.5. ***Questão heavyweight***: *Uma questão heavyweight é aquela que requer uma proposição heavyweight como resposta*[13].

Dretske diria que o "mundo externo existe" é um fato *heavyweight*. Este fato pode ser afirmado pela proposição H_3. Mas suponha que você ignora este fato; isto é, suponha que você não saiba de sua real condição neste mundo; isto é, você está na dúvida se tem um corpo ou se é apenas uma "coisa pensante". Se você passar a se preocupar com isso, você poderá colocar a questão "O mundo externo existe?" para outras pessoas, e esperar pelo melhor. A resposta provavelmente lhe trará tranquilidade. Digo "provavelmente" porque, como você já pode ter percebido, todas as respostas para este tipo de questão (sejam elas positivas ou negativas) serão proposições *heavyweight*. Deste modo, se você não gosta de nada *heavyweight*, você ficará bastante desapontado. Mas o que realmente importa agora é o seguinte:

Proposição 3.6. Conexão heavyweight: *Proposições, fatos e questões heavyweight estão interligados:*

a) Se existem proposições heavyweight verdadeiras, também existem fatos e questões heavyweight;

b) Se existem fatos heavyweight, também existem proposições heavyweight verdadeiras e questões heavyweight;

c) Se existem questões heavyweight, também existem proposições heavyweight (verdadeiras ou falsas) e fatos heavyweight.

Portanto, a aceitação do conceito de *heavyweightness* implica na aceitação da proposição 3.6. A demonstração da conexão existente entre fatos, proposições *heavyweight* e questões *heavyweight* não é difícil de ser percebida, nem demonstrada. Tomemos, por exemplo, este mesmo caso do mundo externo. Suponha que o mundo externo exista. Ora, se ele existe, então há também uma proposição que expressa verdadeiramente este fato, a saber, "O mundo externo existe". Ora, se esta proposição existe, então, por sua vez, existe uma pergunta para a qual a referida proposição é uma resposta bem sucedida; a saber, "O mundo externo existe?".

cimento.

[13]Mesmo que a respectiva resposta seja um simples "sim" (ou um simples "não"), ela pressupõe uma resposta completa. Como exemplo, considere a questão "Sou um prisioneiro da *Matrix*?". A resposta para esta questão pode ser um simples "não". Contudo, por este "não" queremos dizer "Eu não sou um prisioneiro da *Matrix*." – uma resposta completa (e *heavyweight*).

Demonstrações similares podem ser feitas para os itens b) e c)[14].

3.5.2 *Heavyweightness* como incognoscibilidade contingente

Já foi visto no capítulo 1 que a noção de *heavyweightness* – a despeito de ser (in) adequada, (não-)verdadeira etc. – pode ser desenvolvida a partir de recursos muito simples. Talvez seja justamente por isso que ela atraia a atenção de epistemólogos como o próprio Dretske. Afinal de contas, para construir proposições *heavyweight*, precisamos começar apenas com noções muito elementares, noções que dizem respeito às sensações ou experiências do próprio agente que cogita, infere ou decide sobre a verdade ou falsidade de uma dada proposição (ou de um conjunto de proposições). Comecemos, por exemplo, com minhas próprias experiências sensoriais.

Ora, eu sei, através da percepção, que eu tenho mãos: eu posso vê-las, tocá-las, senti-las etc. Entretanto, vimos no primeiro capítulo – na discussão sobre Dretske e as proposições *heavyweight* – que a percepção não é capaz de transmitir garantias evidenciais a proposições que são, por sua vez, consequentes de implicações lógicas nas quais os respectivos antecedentes são proposições com garantias evidenciais. Deste modo, apesar de ver que tenho mãos, eu não vejo que o mundo externo existe (não necessariamente, pelo menos). Ou seja, eu não posso "ver" que o mundo externo existe simplesmente porque também posso ver minhas mãos. Dito de uma maneira mais geral, não posso "ver" que o mundo externo existe com base apenas na informação de que a primeira proposição, a saber, "vejo minhas mãos", implica logicamente a segunda proposição ("o mundo externo existe").

Segundo Dretske, este argumento vale (ou pelo menos deveria valer) para todas proposições (ou implicações) *heavyweight*. Contudo, como bem sabemos, existem outras fontes de conhecimento além da percepção e da razão. Talvez não possamos realmente ter certeza de que o mundo material exista com base apenas em nossa percepção de uma formiga em nossa escrivaninha (isto é, da percepção da formiga). Talvez, de fato, não possamos conhecer proposições *heavyweight* a partir da percepção. Talvez... Mas também é possível que exista uma outra forma de conhecer tais coisas. O próprio Dretske questiona-se acerca dessa possibilidade:

[14]É claro que a demonstração que acabei de fornecer foi apenas particular. Entretanto, se a apresentei apenas como particular, foi tão somente por razões de simplificação. Certamente, não é difícil perceber, pelo significado (neste caso informal) dos termos "perguntas", "respostas" e "fatos", que o mesmos estão interligados, e que a definição de um contribuirá para a definição do outro.

> Deve haver outra forma, além da percepção, na qual eu saiba que não sou um cérebro numa cuba, que eu não estou sendo completamente enganado, que o solipsismo seja falso, que tudo não seja apenas um sonho. Que outras formas de conhecer poderiam ser essas[15]? (DRETSKE, 2005a, p. 20)

Porém, ele continua:

> É difícil ver que outras formas poderia haver, já que cada uma dessas formas de conhecer ou falha em alcançar essas implicações *heavyweight* ou gera suas próprias implicações *heavyweight*. Nenhuma evidência transmite a todas as implicações que é uma evidência[16].

Como se pode observar, a posição de Dretske sobre o fecho do operador de conhecimento é a mesma em relação ao fecho da garantia evidencial. Como sabemos, nós também adquirimos conhecimento através da introspecção, da memória e também através de testemunhos. Porém, segundo Dretske, todas essas fontes de conhecimento não transmitem garantia evidencial para suas implicações lógicas. Assim, eu posso saber – através de um testemunho – que o pneu do meu carro está furado justamente porque alguém de minha confiança acabou de contar-me. Ora, a proposição “O pneu do carro de Stanley está furado” implica nesta outra, a saber, “O mundo externo existe”. A questão, agora, é a seguinte: ser informado de que o pneu do meu carro está furado é o mesmo que ser informado que o mundo externo existe? A resposta de Dretske, é claro, é “não”. Novamente, neste caso, a evidência para a primeira proposição não pode contar como evidência para a segunda. O testemunho não é uma fonte de conhecimento fechada sob implicação lógica. Exemplos similares podem ser construídos para as demais fontes de conhecimento.

Essa argumentação nos leva à tese de que as proposições ditas *heavyweight* não podem ser conhecidas através de quaisquer das fontes de conhecimento que apresentamos (isto e, percepção, razão, memória, introspecção e testemunho). Alguns, como Dretske, iriam mais longe e afirmariam que tais proposições têm justamente essa característica, qual seja, a de não serem epistemicamente acessíveis através de qualquer fonte de conhecimento.

Essa postura frente às proposições *heavyweight* não é, como se pode imaginar, aceita consensualmente. A discussão sobre a cognoscibilidade ou incognos-

[15]*“There must be a way other than perception in which I know that I?m not a brain in a vat, that I?m not being massively deceived, that solipsism is false, that it is not all just a dream. What might these other ways of knowing be?”*

[16]*“It is hard to see what other ways there could be since every way of knowing either fails to reach these heavyweight implications or generates its own heavyweight implications. No evidence transmits to all the implications of what it is evidence for.”* (DRESTKE, 2005a, p. 20)

cibilidade de proposições *heavyweight* ainda é controversa, e é exatamente essa controvérsia que abre espaço para a noção de incognoscibilidade contingente. Ora, uma das possíveis causas desse desacordo acerca da cognoscibilidade de proposições *heavyweight* pode ser justamente a falta de uma qualificação adequada para esse tipo de proposição. Isto é, afirmar ser impossível conhecer proposições como “O mundo externo existe” é assumir o sentido forte da noção de incognoscibilidade, e isso certamente não agradará a todos, dado que tal impossibilidade não fora satisfatoriamente demonstrada. Por outro lado, mesmo aqueles que defendem a cognoscibilidade dessas proposições reconhecem que “O mundo externo existe” não é, por assim dizer, indubitável. Pelo contrário, o próprio Descartes mostrou em sua época que o mundo externo é uma das coisas que podem ser duvidadas, juntamente com inúmeras outras. Assim, a postura dos céticos, ou do próprio Dretske, tem sua razão de ser. Porém, o problema da cognoscibilidade das proposições *heavyweight* ainda persiste sem um consenso.

Essa confusão, ao que parece, está fundamentada nas interpretações diferentes da noção de impossibilidade. Afinal de contas, estamos falando aqui de que tipo de impossibilidade? Para a noção de *heavyweightness*, certamente não poderemos utilizar aqui a noção de impossibilidade lógica; pois, como se pode perceber, a causa da incognoscibilidade de uma proposição *heavyweight* difere e muito daquela que encontramos nas proposições do tipo “Fitch”, isto é, das necessariamente incognoscíveis. Nestas últimas, a incognoscibilidade se dá por razões puramente lógicas, e independem de qualquer contrafactual da contingência. Já as proposições *heavyweight*, como veremos mais adiante, mantêm uma relação “amigável” com esses condicionais. Assim, nossa estratégia será mostrar que a caracterização das proposições “*heavyweight*” como contingentemente incognoscíveis contribuirá para amenizar a tensão existente entre aqueles que são contra ou a favor da incognoscibilidade *simpliciter* das proposições *heavyweight*. Ao meu ver, a estratégia de identificar *heavyweightness* com incognoscibilidade contingente resolve esse impasse na medida em que: (i) preserva a possibilidade lógica de se conhecer proposições *heavyweight*, (ii) permite que se aplique, em certa medida, a noção de incognoscibilidade a proposições *heavyweight* e (iii) não é incoerente com a tese da falha de transmissão de garantia evidencial. Portanto, a aproximação das noções de *heavyweightness* e incognoscibilidade contingente pode ser um caminho viável para o fim da disputa acerca da cognoscibilidade de proposições *heavyweight*. Vejamos, então, como isso ocorre; isto é, como a definição de incognoscibilidade contingente pode ser aplicada à noção de *heavyweightness*.

Com o intuito de aproximar essas duas noções, pensemos, por exemplo, no filme *Matrix*. Para aqueles que aceitam a noção de *heavyweightness*, a hipótese abaixo certamente contaria como uma proposição *heavyweight* para o personagem *Neo*:

H_5: "Eu sou um prisioneiro da *Matrix*."

Muitos hão de concordar que, antes de conhecer *Morpheus*, *Neo* não tinha como conhecer H_5 através de qualquer meio de que ele dispunha na ocasião em que a referida proposição foi cogitada. Isso ocorre porque a informação necessária para a determinação de H_5 não estava disponível a *Neo*. Como sabemos, apenas *Morpheus* e seus colegas poderiam fornecer-lhe essa informação. Agora, observemos que a proposição H_5 asserta algo, a saber, que *Neo* é um prisioneiro da *Matrix* (o que, inclusive, ocorre ser verdadeiro). Assim, H_5 asserta um fato sobre *Neo*, um fato que o próprio Neo ignora plenamente[17]. E não apenas isso: ao que parece, não há nada que *Neo* possa fazer para conhecer a proposição em questão (pelo menos, não sozinho).

Como se pode observar neste caso, em particular, o desconhecimento de *Neo* sobre a verdade de H_5 é algo que lhe é completamente estranho; ou seja, é algo que *Neo*, sozinho, não pode evitar. É plausível admitir que o próprio *Neo*, quando questionado sobre a verdade de H_5, possa chegar a admitir, de sua parte, a incognoscibilidade desta proposição. Em outras palavras, é plausível sustentar que, neste caso, H_5 possa ser considerada uma proposição contingentemente incognoscível para o agente *Neo*. Mais adiante, isso será demonstrado através da aplicação da definição 3.3 a este caso. Por ora, observemos um outro detalhe interessante.

Observamos também que, antes de conhecer *Morpheus*, a pergunta "Sou um prisioneiro da *Matrix*?" seria (contingentemente) irrespondível para *Neo*. Na verdade, temos pelo menos três aspectos importantes da proposição H_5 para considerar, quais sejam:

(1) *Neo* não conhece a proposição H_5.

(2) Que *Neo* é um prisioneiro da *Matrix* é um fato, e *Neo* não sabe disso; as evidên-

[17]Note-se que o não ter informação suficiente é uma causa contingente. Se as coisas tivessem sido diferentes, isto é, se a requerida informação fosse fornecida, o agente em questão (*Neo*) conheceria a respectiva proposição.

cias para conhecer a respectiva proposição não estão disponíveis para *Neo*.

(3) "Sou um prisioneiro da *Matrix*?" é uma pergunta (contingentemente) irrespondível para *Neo*.

Agora, simplesmente juntemos (1), (2) e (3) e o seguinte contrafactual:

MC: Se *Morpheus* contasse a *Neo* sua verdadeira situação, então *Neo* saberia que ele era um prisioneiro da *Matrix*[18].

Agora, além de MC, consideremos também a seguinte proposição:

P_m: "*Morpheus* contou a *Neo* sobre sua verdadeira situação".

A adição de MC e P_m a (1)-(3) nos traz o seguinte resultado:

Se Pm e MC forem ambos satisfeitos (verdadeiros), então, como resultado, teremos:

(1a) *Neo* conhece a proposição H_5.

(2a) Que *Neo* é um prisioneiro da *matrix* é um fato, e ele sabe disso.

(3a) "Sou um prisioneiro da *Matrix*?" foi corretamente respondida por *Neo*.

Observamos, portanto, que a ajuda de *Morpheus* muda o status da proposição H_5. Antes de *Morpheus*, H_5 era epistemicamente inacessível a *Neo*. Porém, como podemos observar acima, essa inacessibilidade (ou limitação, se preferir) não era de caráter lógico, mas tão somente contingente. Isto é, H_5 era contingentemente inacessível a *Neo*, mas não "em princípio". Isso significa dizer que a possibilidade lógica de Neo conhecê-la permaneceu aberta, até que condições especiais (entre elas um contrafactual da contingência) foram satisfeitas, e H_5 mudou o status de "contingentemente inacessível" para "epistemicamente acessada". Como já argumentamos, essa é, justamente, uma das características das proposições contingentemente incognoscíveis.

A partir de agora, será fácil compreender porque a definição de incognoscibilidade contingente pode, com sucesso, ser aplicada a este caso da *Matrix*. Seja H_5 a proposição em questão, *Neo* o agente em questão e MC o contrafactual em questão. Temos, portanto, o seguinte:

[18]Em outras palavras, se *Morpheus* contasse a *Neo* sua verdadeira situação, ele conheceria a proposição H_5.

1. *Neo* não conhece a proposição H_5.

2. H_5 é, neste contexto, uma proposição que carece de evidências de suporte e, além disso, a probabilidade de se encontrar tais evidências, no momento t (que *Neo* cogitou a possibilidade de ser um prisioneiro da *Matrix*), é considerada muito baixa.

3. MC é um contrafactual da contingência para H_5.

Observe-se que o "caso da *Matrix*" satisfaz as três condições da definição de incognoscibilidade contingente. Facilmente, podemos concordar com a satisfação das condições 1 e 3. Afinal de contas, uma das características marcantes do filme *Matrix* é o fato de seu protagonista, *Neo*, no início do filme, desconhecer sua verdadeira condição de prisioneiro da *Matrix*. Em seguida, observamos que MC não somente desempenha a função de contrafactual da contingência para a proposição H_5, mas também chega a ser satisfeito no decorrer da trama, alterando o *status* da proposição H_5. A condição possivelmente problemática, entretanto, é a segunda.

A condição original, isto é, aquela encontrada na definição 3.3, exige que H_5 seja "consensualmente considerada, no contexto conversacional em uso – isto é, em que S (*Neo*) está inserido – uma proposição que carece de evidências de suporte e, além disso, a probabilidade de se encontrar tais evidências, no momento t (em que *Neo* cogita sobre a *Matrix*), seja considerada muito baixa". Obviamente, a questão agora deve naturalmente voltar-se "àqueles que estão em consenso". Quem são eles? Comumente, esses elementos são chamados de "atribuidores de conhecimento"[19], e variam segundo o contexto.

Certamente, a questão final de saber se a proposição H_5 é ou não acessível epistemicamente a *Neo* dependerá de quem atribui, no atual contexto, conhecimento ao agente. No presente caso, nós, que consideramos o filme como espectadores, facilmente somos levados a admitir que, nas circunstâncias em que se encontra, o agente *Neo* é completamente incapaz de vir a conhecer (sozinho, e no momento t) a proposição H_5. Somos levados consensualmente – enquanto espectadores que assistem e entendem a trama – a reconhecer que a proposição H_5 é epistemicamente inacessível ao agente *Neo*, a menos que certos condicionais contrafactuais sejam satisfeitos. Em outras palavras, concordamos com a falta de evidência do agente *Neo* em relação à referida proposição; concordamos acerca da baixa proba-

[19]Em inglês, *knowledge ascribers*.

bilidade do agente adquirir – no momento em que cogita a proposição em questão – evidências que suportem seu conhecimento nela; além disso, enquanto espectadores que compreendem a trama, temos esperanças que o contrafactual MC seja satisfeito. Para isso, reconhecemos que um conjunto de condições deve ser satisfeito (Neo escolher a pílula vermelha etc.), condições sem as quais o agente Neo não seria capaz de conhecer sua verdadeira condição de prisioneiro da *Matrix* etc.

Colocando de outro modo, o próprio sucesso na captação da ideia central do filme *Matrix* exige que os espectadores desta obra compreendam e aceitem três coisas:

(1) *Neo* não sabia que era um prisioneiro da *Matrix*.

(2) *Neo* não tinha como saber que era um prisioneiro da *Matrix*, no momento em que passou a considerar essa hipótese; ele não dispunha de quaisquer recursos que pudessem ajudá-lo a descobrir, sozinho, sua verdadeira condição. A probabilidade de conhecer, por si mesmo, que ele era um prisioneiro da *Matrix* era muito baixa.

(3) Se *Morpheus* o ajudasse, *Neo* saberia que era um prisioneiro da *Matrix*. Por outro lado, se *Morpheus* não o fizesse, *Neo* nunca saberia acerca de sua condição de prisioneiro da *Matrix*. Ou, se *Neo* tivesse escolhido a pílula vermelha, ele saberia que era um prisioneiro da *Matrix*; por outro lado, se tivesse escolhido a pílula azul, ele não saberia que era um prisioneiro da *Matrix*.

Assim, a admissão da proposição H_5 como contingentemente incognoscível parece ser uma exigência feita aos espectadores desta obra, exigência esta bastante sutil.

Observa-se, tal como foi demonstrado, que H_5 satisfaz plenamente todos os pré-requisitos de uma proposição contingentemente incognoscível. Deste modo, a definição 3.3 pode, no caso acima, ser aplicada com sucesso. Como conclusão parcial, portanto, sustentamos a caracterização da proposição H_5 como contingentemente incognoscível. Agora, generalizando o resultado – já que a proposição *heavyweight* H_5 foi escolhida arbitrariamente – conclui-se que todas as proposições *heavyweight* podem, pela aplicação da definição 3.3, ser caracterizadas como proposições contingentemente incognoscíveis. Isso ocorre porque, para qualquer proposição P que seja aceita como *heavyweight* por um determinado agente, S, as seguintes condições podem ser satisfeitas:

1. S não conhece P;

2. É consensual que P carece de evidências, e que a probabilidade de S adquirir tais evidências é muito baixa[20];

3. Existe um contrafactual da contingência para P.

É claro que isso não significa, por exemplo, que todos devem aceitar a incognoscibilidade *simpliciter* de proposições como "Não sou um prisioneiro da *Matrix*" ou "Não sou um cérebro numa cuba". Ou seja, não está sendo afirmado, neste argumento, que não podemos saber se somos ou não cérebros em cubas, marionetes do gênio maligno cartesiano ou prisioneiros da *Matrix*. Ao invés disso, o que está sendo afirmado é o seguinte:

Proposição 3.7. (*heavyweightness* – incognoscibilidade contingente) Se P é uma proposição *heavyweight* para um agente qualquer S, então P é contingentemente incognoscível para o agente S.

Ou seja, se P já é considerada *heavyweight* para dado agente (ou por um dado agente, no caso de P ser considerada pelo próprio agente em questão), P será contingentemente incognoscível para esse agente. Entretanto, como sabemos, isso não significa afirmar que é logicamente impossível para S conhecer P: as razões que tornam P incognoscíveis são, como bem sabemos, apenas contingentes, e não lógicas ou necessárias.

A estratégia de caracterização das proposições *heavyweight* como contingentemente incognoscíveis atende bem, ao meu ver, às exigências tanto daqueles que defendem a incognoscibilidade de proposições *heavyweight* (DRETSKE, 2005a e 2005b) quanto daqueles que defendem exatamente o oposto, a saber, que é possível conhecer proposições tidas como *heavyweight* (HAWTHORNE, 2005). Por um lado, as proposições contingentemente incognoscíveis não podem ser conhecidas senão pela satisfação de condições especiais, de contrafactuais especialmente adequados a elas (e de difícil satisfação). Por outro lado, esta estratégia mostra o quanto é inadequado confundir a incognoscibilidade das proposições *heavyweight* com a incognoscibilidade lógica (e necessária) de algumas proposições que aparecem nos teoremas de Fitch (1963). Ou seja, a estratégia em questão demonstra a possibilidade lógica da cognoscibilidade de proposições *heavyweight*. Creio, portanto, que a identificação da noção de *heavyweightness* com a noção de incognosci-

[20] O agente epistêmico, mesmo quando sozinho, pode pensar sobre P e chegar à conclusão de que não tem evidências suficientes para sustentar o conhecimento em P.

bilidade contingente é um caminho interessante e que contribui para a redução da tensão existente acerca do conceito de *heavyweightness*, e também contribui positivamente para os trabalhos de análise de alguns princípios de fecho epistêmico – esse resultado, porém, ficará claro apenas no final deste capítulo.

3.6 Hipóteses céticas e incognoscibilidade contingente

Já podemos perceber, pelo que vimos até agora, que a definição de incognoscibilidade contingente também pode ser aplicada às tão conhecidas "hipóteses céticas". O procedimento será, basicamente, o mesmo adotado anteriormente. Alguns cuidados, porém, devem ser tomados, já que hipóteses céticas e proposições *heavyweight* não são exatamente a mesma coisa.

Uma hipótese cética como, por exemplo, "sou um cérebro numa cuba", tem o propósito cético de questionar a certeza de algum conhecimento que esteja direta ou indiretamente relacionado com ela (através, provavelmente, de uma negação). Um exemplo imediato disso poderia ser "não sou um cérebro numa cuba". Esta é uma proposição diretamente relacionada à hipótese cética "sou um cérebro numa cuba", e essa relação se dá através do operador de negação e tudo aquilo que ele representa. Além disso, como exemplo indireto, poderíamos pensar em "Stanley se saiu bem em sua defesa de dissertação, em 2008". Considerando que cérebros não defendem dissertações, essa proposição só poderia ser verdadeira se "Stanley é um cérebro num cuba" fosse falsa ou, de modo análogo, "Stanley não é um cérebro numa cuba" fosse verdadeira. Assim, hipóteses céticas têm geralmente o objetivo de questionar o conhecimento de outras proposições, e é de se esperar que o operador de negação esteja presente na relação hipótese cética/proposição questionada pela hipótese cética.

Com as proposições *heavyweight*, as coisas nem sempre são assim. Tomemos como exemplo a proposição "o mundo externo existe". Esta proposição não tem, pelo menos em princípio, o objetivo de questionar o conhecimento de alguma proposição ou fato qualquer. Temos aqui um afirmação positiva, que não precisa estar relacionada necessariamente a uma hipótese cética. É claro que, se quisermos, podemos associá-la a proposições como "o ceticismo está incorreto" ou "o solipsismo é falso" etc. No entanto, como podemos perceber, apesar de hipóteses céticas satisfazerem a definição de *heavyweightness*, as proposições *heavyweight* não precisam ser hipóteses céticas. Em suma, todas as hipóteses céticas são *heavyweight*, mas nem todas as proposições *heavyweight* são hipóteses céticas. Entretanto, ambas, proposições *heavyweight* e hipóteses céticas, satisfazem a noção de incognoscibilidade contingente. Parte disso já foi demonstrado na seção anterior, em que tomamos as proposições *heavyweight* como contingentemente incognoscíveis. No que se segue, demonstraremos com poucas palavras que o mesmo se aplica às hipóteses céticas. Na verdade, essa demostração é imediata se reconhecermos que a proposição *heavyweight* do exemplo anterior, a saber, H_5 = "Sou um prisioneiro da *Matrix*", também pode ser considerada uma hipótese cética. Ora, não é preciso muito esforço para aceitar que H_5 é a negação direta da proposição "Não sou um prisioneiro da *Matrix*"; ou seja, H_5 pode ser considerada como uma hipótese cética que questiona o conhecimento daquilo que é expresso pela proposição "Não sou um prisioneiro da *Matrix*", que é justamente o fato de não sermos prisioneiros de realidades virtuais ou coisa do gênero. Deste modo, todas as considerações da seção anterior também se aplicam aqui. Isto é, hipóteses céticas também podem ser caracterizadas como proposições contingentemente incognoscíveis.

Entretanto, isso não quer dizer que a afirmação não precisa ser demonstrada com igual rigor; sua demonstração segue o mesmo método, contanto que sejam feitas as devidas adaptações às cláusulas 1 e 2:

1. Eu não sei que não sou um prisioneiro da *Matrix*;

2. É consensual que "Eu não sou um prisioneiro da *Matrix*." carece de evidências, e que a probabilidade de eu adquirir tais evidências é muito baixa.

Ora, a aceitabilidade destas cláusulas vai depender do que penso sobre o assunto, juntamente o que pensam aqueles que estão considerando as mesmas coisas que eu. De fato, pode haver um desacordo sobre a (in) incognoscibilidade da proposição em questão. No entanto, como já foi frisado, isso não é o que está em jogo no momento. O que observamos agora é que é perfeitamente plausível que exista um conjunto de agentes consideradores de hipóteses céticas (eu, Dretske e seus defensores, por exemplo) que estejam em consenso tanto sobre "1" quanto sobre "2". Assim, se insisto em manter que não sei se sou ou não um prisioneiro da *Matrix*, posso perfeitamente aplicar a noção de incognoscibilidade contingente e chegar à conclusão de que a proposição em questão é contingentemente incognoscível. Isso ocorre porque ela satisfaz tanto as cláusulas 1 e 2, quanto a cláusula 3 (que é facilmente formada por um condicional contrafactual específico para este caso). Portanto, novamente, hipóteses céticas podem ser caracterizadas como proposições contingentemente incognoscíveis. Isso não quer dizer, é claro, que elas devam ser consideradas apenas deste modo. Para um conjunto diferente de agentes ou uma aplicação diferente (ou contexto, se preferir), a cognoscibilidade de H_5 pode ser algo natural. Todavia, tudo o que o argumento sugere aqui é que, se há um consenso sobre a incognoscibilidade de H_5 entre um grupo finito de agentes, o modelo da incognoscibilidade contingente oferece uma caracterização dessa incognoscibilidade e estará à disposição sempre que precisarmos dele.

Para aqueles, como Dretske, que rejeitam o fecho epistêmico devido à existência de proposições *heavyweight* – que não ganharam uma caracterização clara deste então, para dizer a verdade – o modelo da incognoscibilidade contingente pode ser útil na análise da relação "princípios de fecho/agentes consideradores de hipóteses céticas". É justamente essa relação, com base nas noções de incognoscibilidade necessária e contingente, que esta investigação pretende elucidar. Faremos isto a seguir.

3.7 Princípios de fecho e incognoscibilidade contingente

No capítulo anterior, concluímos que a simplificação exagerada da discussão sobre os princípios de fecho impede uma compreensão mais profunda desses princípios, de modo que o problema da validade/invalidade de um princípio de fecho é algo que deve ser considerado da mesma forma que na epistemologia formal, isto é, com a restrição adequada do conceito de validade à lógica de interesse, bem como sua pretensão de modelagem. Assim, um princípio de fecho será aceitável ou inaceitável em virtude da situação que ele pretende modelar. Falar em validade ou invalidade *simpliciter* de um determinado princípio de fecho constitui um erro de super-simplificação.

Nesta seção, seguindo esta estratégia da epistemologia formal, analisaremos os princípios de fecho na aplicação de interesse deste trabalho de investigação, isto é, na aplicação em que determinados agentes consideram se conhecem ou não proposições que constituem hipóteses céticas. Como mostramos na seção anterior, tais proposições, juntamente com as proposições *heavyweight* de Dretske, podem ganhar o status de “incognoscíveis”; mais especificamente, podem ser caracterizadas como proposições contingentemente incognoscíveis.

Assim, considerando hipóteses céticas como “sou um cérebro numa cuba” ou “sou uma marionete do gênio maligno” como proposições contingentemente incognoscíveis para um dado número de agentes, os fechos epistêmicos que consideramos no capítulo anterior, a saber, (E-CLOS 1), (E-CLOS 2), (E-CLOS 3) e (E-CLOS 6) são todos inválidos ou inaceitáveis para esta aplicação. Em suma, os princípios mencionados acima são considerados não-aplicáveis às situações específicas em que agentes consideram seriamente hipóteses céticas – que neste contexto passam a ser caracterizadas como proposições contingentemente incognoscíveis por aqueles que atribuem conhecimento aos agentes (podendo estes serem os próprios agentes em questão). Comecemos com o principal deles, a saber, (E-CLOS 1):

(E-CLOS 1): Se S conhece P e conhece $(P \to \neg Q)$, então S conhece $\neg Q$.

Seja P a proposição “Estou escrevendo uma tese em epistemologia” e $\neg Q$ a proposição “Não sou um cérebro numa cuba”. Neste momento, considero se conheço ou não a proposição $\neg Q$. Até agora, sei que, nesta aplicação, conheço P (estou escrevendo agora mesmo) e também conheço o fato de P implicar $\neg Q$, ou $\neg Q$ ser uma consequência lógica de P etc. Agora, se $\neg Q$ for aceita por mim como contingentemente incognoscível (e assim a vejo, assim como todas as hipóteses céticas), reconhecerei que não conheço e nem tenho como conhecer $\neg Q$. Para isso, basta que reconheça aquilo já exigido na definição de incognoscibilidade contingente:

1. Não sei se sou ou não um cérebro numa cuba;
2. Reconheço que “Não sou um cérebro numa cuba” carece de evidências suficientes, e que a probabilidade de eu adquirir tais evidências é muito baixa.
3. (C-FACT) é um contrafactual da contingência para $\neg Q$[21].

Assim, nesta situação, o princípio (E-CLOS 1) é inaplicável, pois observamos que o agente em questão, isto é, eu, conhece P e conhece $(P \to \neg Q)$, mas não conhece $\neg Q$, pois esta é considerada uma proposição contingentemente incognoscível. É claro que este resultado não precisa se repetir para todas as situações

[21]Pense, neste caso, (C-FACT) como qualquer contrafactual tal que, se fosse satisfeito, eu conheceria $\neg Q$. Exemplo: “Se Deus existisse e quisesse tirar de Stanley suas dúvidas céticas e Ele, enquanto onipotente, quisesse ajudar Stanley e o mostrasse a verdade sobre sua condição neste mundo, então Stanley saberia não ser um cérebro numa cuba” etc.

similares. O que está em jogo aqui é o status epistêmico das proposições céticas. Outros atribuidores de conhecimento poderiam, neste caso, discordar do fato de $\neg Q$ ser contingentemente incognoscível e, assim, defenderem a aplicação bem sucedida de (E-CLOS 1). Todavia, do mesmo modo que aceito esta possibilidade, insisto também na aceitação da possibilidade de que um agente, ou um grupo de atribuidores de conhecimento a agentes, reconheça que as hipóteses céticas satisfazem a condição de proposições contingentemente incognoscíveis e que, nesta perspectiva, princípios de fecho epistêmico que as envolvam não sejam aplicáveis.

Logo, apesar de seguir uma estratégia um pouco diferente da de Dretske, concluo que o princípio (E-CLOS 1) é inválido ou, melhor dizendo (para evitar super-simplificação), que ele não se aplica a algumas situações. Tais situações são, como mostrei, aquelas em que atribuidores de conhecimento tomam hipóteses céticas como proposições contingentemente incognoscíveis. A grande diferença entre Dretske e o argumento que ofereço aqui é que não estou simplesmente afirmando a invalidade do fecho (E-CLOS 1). Estou sugerindo que, com as pretensões de aplicação que tenho em mente – isto é, modelar agentes que consideram aquilo que chamei de "proposições contingentemente incognoscíveis" – o princípio (E-CLOS 1) não me interessa e, portanto, pode ser descartado nessas situações específicas. Isto não significa afirmar a invalidade geral deste princípio mas tão somente que, para a lógica e para a aplicação que tenho em mente (que é a de modelar agentes consideradores de proposições contingentemente incognoscíveis), ele não me é necessário nem aplicável. Assim, temos um resultado que é coerente com aquele oferecido por Dretske, porém um pouco mais cauteloso – na medida em que leva em consideração os desenvolvimentos da epistemologia formal sobre o tema, e adota a mesma estratégia de análise de princípios de fecho – isto é, na perspectiva das situações de modelagem de interesse e da aplicação. Logo, certos princípios de fecho não são aplicáveis e todas as situações, mas isso não significa afirmar que são inválidos, de um modo geral – pois, assim como na epistemologia formal, a noção de validade é particular a cada lógica e, portanto, perguntar pela validade ou invalidade de uma fórmula arbitrária é perguntar por sua validade dentro de uma lógica previamente definida.

Todas as considerações para o princípio (E-CLOS 1) se aplicam aos demais princípios já mencionados. Colocando (E-CLOS 2) numa "roupagem" mais informal, temos mais ou menos o seguinte:

(E-CLOS 2)* Se S conhece um conjunto de sentenças X e X implica logicamente a proposição Q, então S conhece Q.

Novamente, se Q for uma proposição contingentemente incognoscível, ela é, por definição, contingentemente e epistemicamente inacessível ao agente S, mesmo que ela seja uma consequência lógica do conjunto X (de proposições que S conhece). Assim, nesta situação, o princípio (E-CLOS 2)* igualmente ou não se aplica ou sua aceitação não é necessária. Além disso, como vimos no capítulo anterior, (E-CLOS 2)* pode não ser aceito por uma razão bem mais simples: ele não explicita que o agente em questão conhece a implicação relevante $(X \rightarrow P)$. Assim, S pode falhar em conhecer P justamente por não conhecer a implicação $(X \rightarrow P)$. Aqui, vemos claramente que o mesmo se aplica à versão informal do princípio (E-CLOS 6):

(E-CLOS 6)* Se S conhece P e $(P \rightarrow Q)$ é uma verdade, então S conhece Q[22].

De modo análogo ao caso anterior, Q pode ser uma proposição contingentemente incognoscível e por isso, por definição, não pode ser conhecida; ou pode ser o caso que o agente desconheça a implicação $(P \rightarrow Q)$ que, neste caso, poderia ser relevante para a determinação do conhecimento de S na proposição Q. A versão epistemológica do princípio (E-CLOS 3), dada no primeiro capítulo, é a seguinte:

(M-CLOS 3) (conhecimento de verdades): Se t é uma verdade, então o agente S conhece t.

Aqui, não há nada que proíba t de ser uma verdade lógica (como vimos no primeiro capítulo) mas contingentemente incognoscível. No início deste capítulo, vimos que não podemos fornecer exemplos de proposições incognoscíveis a nós mesmos: o próprio ato de tentar fornecer um exemplo de uma proposição incognoscível para nós destrói aquilo que está sendo exigido, isto é, que esta última seja

[22] A expressão "$(P \rightarrow Q)$ é uma verdade" é, sem dúvida, um tanto estranha. Todavia, é comum encontrarmos em Rescher (2005, p. 68 e 2010, p. 10) expressões como "verdade incognoscível" ("*unknowable truth*") e "verdade não-especificável" ("*unspecifiable truth*") etc. Em outra ocasião (RESCHER, 2010, p. 67), ele associa a palavra "verdade" ("*truth*") à palavra "fato" ("*fact*"): "*being a truth (a fact) no one has ever realized (learned, stated)*" – isto é "ser uma verdade (fato) que ninguém nunca se deu conta (aprendeu, afirmou)". Talvez fosse mais indicado, para evitar esse tipo de confusão, utilizar a distinção entre proposições "logicamente verdadeiras" e "contingentemente verdadeiras". A ideia, porém, de usar a expressão "uma verdade" é simplesmente para deixar a aplicação do princípio em questão mais genérica. Assim, ele estaria referindo-se tanto a proposições da primeira categoria quanto da segunda. A utilização da expressão "uma verdade" é especialmente importante para o princípio (E-CLOS 3): a ideia é que ele refira-se não apenas a verdades lógicas, mas também a fatos contingentes e, portanto, a proposições que sejam contingentemente verdadeiras. De qualquer modo, a formulação atual será mantida para que não nos distanciemos demais dos termos empregados por Rescher em suas investigações sobre incognoscibilidade.

incognoscível para nós. Porém, vimos igualmente que isso não impede que proposições dessa categoria existam. Assim, neste caso em particular, t pode constituir justamente uma verdade logica contingentemente incognoscível para nós. Assim, mesmo que seja uma verdade lógica, não é o caso que conheçamos P. Portanto, para esta situação, este princípio de fecho não é aplicável. Aqui, especificamente, observa-se que não é difícil questionar a plausibilidade de (E-CLOS 3), mesmo sem a utilização de proposições contingentemente incognoscíveis. Como não somos logicamente oniscientes e não conhecemos todas as verdades lógicas, existe ao menos uma proposição logicamente verdadeira que não conhecemos. Assim, demonstramos que os princípios de fecho epistêmico acima expostos não se aplicam a todos os casos. Mais especificamente, considerando situações em que um agente, ou grupo de agentes, considera seriamente hipóteses céticas – caracterizadas aqui como proposições contingentemente incognoscíveis – tais princípios de fecho epistêmico não se aplicam.

3.8 Considerações sobre os resultados

Em suma, demonstramos que a noção de incognoscibilidade contingente pode ser utilizada com sucesso em situações em que os atribuidores de conhecimento estão em consenso sobre a incognoscibilidade de hipóteses céticas. Esta característica de serem apenas contingentemente incognoscíveis, típica de proposições céticas, é o que as diferencia das proposições logicamente incognoscíveis do tipo Fitch, como vimos em seções anteriores. Deste modo, oferecemos uma abordagem que define as proposições contingentemente incognoscíveis e fornece uma explicação de porque certos princípios de fecho não são aplicáveis quando estão relacionados a esse tipo de proposição. Tal abordagem se mantém coerente com a tese de que princípios como (E-CLOS 1) não são válidos *simpliciter*, como querem alguns epistemólogos. Porém, ao mesmo tempo, demonstra também que os desenvolvimentos da lógica epistêmica ou epistemologia formal na busca de solucionar o problema da onisciência lógica (capítulo 2) levou a qualificarmos melhor a questão da validade/invalidade de princípios de fecho epistêmico, de modo que a mera busca pela validade/invalidade de um determinado tipo de fecho, sem mais especificações da lógica a ser utilizada e de sua pretensão de modelagem, constitui uma simplificação exagerada dos problemas lógicos e epistemológicos gerados por certos tipos de princípios de fecho. Deste modo, a problemática do fecho epistêmico na perspectiva epistemológica informal deve levar em conta os resultados da epistemologia formal sobre o tema, e considerar todas as tentativas lógicas de solução de problemas relacionados a princípios de fecho. Se assim proceder, a epistemologia *mainstream* acabará por aceitar que o problema do fecho epistêmico, informalmente falando, se desdobrará em uma séries de problemas particulares, todos eles dependentes de lógicas particulares e suas respectivas intenções de modelagem e aplicação.

Conclusão

No primeiro capítulo, mostramos que a discussão sobre o fecho epistêmico, no âmbito epistemológico informal da epistemologia *mainstream*, não chegou ao fim. Afinal de contas, aquilo que ficou conhecido como "o desafio de Dretske" (acerca da invalidade do fecho epistêmico para o operador de conhecimento), até agora, não foi plenamente solucionado. E não há qualquer razão forte para acreditarmos que venha a ser, se continuarmos adotando sempre a mesma estratégia: a construção de contra-exemplos para os mais variados princípios de fecho epistêmico, sem uma especificação clara das intenções de modelagem e aplicação desses princípios.

Ao que parece, a discussão "ingênua" sobre a validade de princípios de fecho tem levado sempre aos mesmos resultados: uns constroem contra-exemplos inteligentes e bem elaborados para um ou outro princípio de fecho (Dretske e de Almeida são exemplos; o primeiro, utilizando como pressuposto proposições *heavyweight*; já o último, utilizando apenas noções simples como "equivalência", da lógica elementar). Outros, por sua vez, elaboram e defendem suas próprias versões de principio de fecho, e tentam sustentá-las como válidas, mostrando que elas sobrevivem a tais e tais contra-exemplos (Hawthorne é um desses).

Da maneira como a discussão sobre os princípios de fecho, no plano epistemológico informal, vem sendo desenvolvida, não há como saber quando, ou até mesmo se, teremos uma solução satisfatória para o problema. Primeiro porque, para uma dada versão de fecho epistêmico, digamos, X, sempre há discordância se X resiste ou não a supostos contraexemplos; além disso, não há como saber se esta mesma versão de fecho, a saber, X, ainda que resista a tais contraexemplos, vai ou não resistir a futuros contraexemplos, diferentes daqueles que estão à disposição atualmente. Além disso, mesmo que seja concebível a construção de contraexemplos que invalidem todas as versões conhecidas (atuais) de fecho epistêmico, isso não garante que uma nova versão seja elaborada, e que esta, por sua vez, seja imune a todas as tentativas conhecidas de invalidação. A dificuldade gerada pela epistemologia *mainstream*, portanto, é simplificar exageradamente a discussão sobre os princípios de fecho epistêmico, ao fazer perguntas como:

1. O princípio de fecho "X" é válido ou inválido?

2. A nova versão modificada de "X" (para evitar certos contraexemplos conhecidos), é válida ou inválida?

3. O operador de conhecimento é fechado sob implicação?

Todas essas perguntas tomam como algo dado, e não levam em consideração a pretensão de modelagem da lógica epistêmica subjacente à análise do princípio de fecho em foco; além disso, não especificam a aplicação desejada para este princípio, além de desconsiderarem o fato de a noção de validade não ser padrão em todas as lógicas; isto é, a própria noção de validade, que é semântica, é definida dentro de cada sistema, e nem sempre vale "fora" do sistema para o qual foi desenvolvida. Ou seja, o que é válido em um sistema de lógica epistêmica pode não ser em outro, e vice-versa. Suponha, por exemplo, que φ é uma fórmula válida de um sistema S qualquer. Assim, dizer que φ é válida é dizer que φ é válida em S, pois φ é uma fórmula bem formada de S. Não faria sentido tirar φ de seu contexto de aplicação e perguntar sobre sua validade: para saber a validade de φ, precisamos sempre associá-la a um sistema específico, que possui uma definição rigorosa (e específica) de fórmula bem formada e validade. Deste modo, na epistemologia formal, tal como não faz sentido perguntar se a fórmula epistêmica $K_a(P_a \to Q_a) \to (K_aP \to K_aQ)$ é (in)válida *simpliciter*[1], igualmente não faz sentido, no âmbito da epistemologia informal, perguntar sobre a (in)validade *simpliciter* de um princípio de fecho "X" qualquer.

Este resultado é uma consequência inevitável da observação da estratégia adotada pela epistemologia formal. No segundo capítulo, investigamos o famoso problema da onisciência lógica, que consiste na aceitação ou rejeição de certos princípios de fecho em sistemas formais de lógica epistêmica – justamente devido à propriedade de "onisciência lógica", considerada inadequada em algumas aplicações.

Vimos que existem diversos princípios de fecho e diversas lógicas epistêmicas diferentes, cada uma capaz de modelar um aspecto particular de falha de onisciência lógica. Conseguimos mostrar que, se a pretensão de modelagem das lógicas estudadas visar a captação de todos os motivos de falha de onisciência lógica,

[1]Isto é: (i) desconsiderando o fato de que a noção de validade é algo definível dentro de um sistema específico; (ii) não especificando que tipo de agentes epistêmicos a lógica utilizada pretende captar; (iii) não especificando a aplicação pretendida para o princípio em questão etc.

então nenhuma delas se mostra adequada. Em contrapartida, todas são bem sucedidas em representar um ou outro aspecto da falha de onisciência lógica. Nessa perspectiva, todas as lógicas estudadas são modelos utilizáveis, se a intenção de modelagem e aplicação for compatível com as capacidades de representação de cada uma dessas lógicas. Seja como for, observamos que algumas dessas lógicas, ainda que sejam similares em alguns pontos, lidam com as mesmas fórmulas (que representam princípios de fecho) de modo diferente. Na lógica epistêmica clássica desenvolvida por Hintikka, por exemplo, é possível demonstrar as propriedades de "onisciência lógica total" (E-CLOS 2), "fecho sob implicação material" (E-CLOS 1), "fecho sob implicação válida" (E-CLOS 6) e "conhecimento de fórmulas válidas" (E-CLOS 3). Já numa versão modificada desta mesma lógica, que considera "impossíveis mundos possíveis", podemos demonstrar a invalidade de (E-CLOS 6), isto é, do fecho sob implicação válida. Deste modo, a versão modificada da lógica de Hintikka, que considera "impossíveis mundos possíveis", é capaz de captar a falha de onisciência lógica por razões de "atenção desconexa", mas não as demais.

A lógica de Levesque, por outro lado, faz distinção entre "crenças implícitas e crenças explícitas", e lida com os mesmos princípios de modo diferente. O fecho sob implicação, por exemplo, vale apenas para as crenças implícitas, e o mesmo pode ser aplicado ao conhecimento de fórmulas válidas e ao fecho sob implicação válida. Assim, nem um desses princípios é aplicável às crenças explícitas. O que observamos, a partir desses e de outros resultados, foi que a lógica das crenças explícitas e implícitas de Lesveque é apropriada para captar a falha de onisciência lógica a partir da falha em "estar ciente" e também da "atenção desconexa". Para os demais motivos, entretanto, essa lógica não se mostra adequada.

A lógica da consciência, apesar de igualmente fazer distinção entre crenças implícitas e crenças explícitas, mantém alguns resultados diferentes daqueles encontrados na lógica de Levesque. O fecho das crenças explícitas, que na lógica de Levesque é inválido, passa a ser válido na lógica da consciência. Já a lógica da consciência geral, que introduz o operador sintático A (de *awareness*), mantém resultados diversos daqueles encontrados na lógica da consciência. Nesta lógica, a crença explícita em uma proposição P, por exemplo, é a conjunção entre uma crença implícita em P e o ato de estar ciente de P – isto é, a pretensão de modelar as crenças explícitas inclui a noção de "consciência dos conceitos relevantes". Desta forma, diferentemente da lógica anterior, a lógica da consciência geral sustenta o fecho sob implicação válida. No entanto, é possível adaptar essa lógica para vários propósitos, colocando-a para validar ou invalidar teoremas segundo a conveniência de aplicação – isso é feito através de imposição de restrições ao operador de consciência, que é essencialmente sintático. Isso nos leva às abordagens sentenciais, que são capazes de invalidar todos os princípios de fecho apresentados, ou de sustentá-los todos, segundo as intenções de aplicação. Novamente, o que fica claro com esses resultados é que o modo como certos princípios de fecho são abordados depende da lógica que os investiga, segundo sua intenção de modelagem e aplicação. Na epistemologia formal, nunca se deve considerar se um determinado princípio de fecho (representado através de uma fórmula epistêmica) é válido ou inválido, sem antes se especificar que lógica está sendo utilizada, e que motivo de falha de onisciência lógica ela pretende captar. A epistemologia informal, por sua vez, deve levar as mesmas coisas em consideração, e evitar simplificações exageradas ao reduzir princípios de fecho a "validos" ou "inválidos" *simpliciter*. De fato, considerando que os princípios de fecho estudados por ambas as epistemologias são exatamente os mesmos; considerando que ambas as epistemologias complementam uma a outra, não há razão para desconsiderar os avanços da epistemologia formal neste tema; não há razão para deixar de adotar a mesma estratégia na análise de princípios de fecho: segundo as intenções de modelagem e aplicação.

Seguindo esta linha de raciocínio, isto nos leva aos resultados do terceiro capítulo: a não-aplicabilidade dos princípios de fecho "onisciência lógica total" (E-CLOS 2), "fecho sob implicação material" (E-CLOS 1), "fecho sob implicação válida" (E-CLOS 6) e "conhecimento de fórmulas válidas" (E-CLOS 3) a situações particulares em que os agentes epistêmicos são conjecturadores de hipóteses céticas. Isso ocorre na medida em que as hipóteses céticas, ou algumas proposições *heavyweight*, são consideradas por estes agentes "proposições contingentemente incognoscíveis" – pois sua incognoscibilidade, apesar de ser reconhecida por estes últimos, se dá por razões puramente contingentes, e isso fica evidente na medida em que as contrastamos com as proposições "necessariamente incognoscíveis" do tipo Fitch, que são incognoscíveis por razões lógicas ou demostrativas.

Em suma, o problema do fecho epistêmico, aparentemente sem uma solução adequada na epistemologia *mainstream* (ou informal), deve ser investigado a partir da estratégia da epistemologia formal, a saber, da análise de princípios de fecho segundo intenções de modelagem e aplicação – tal como vimos com relação ao problema da onisciência lógica. Nessa perspectiva, é possível construirmos uma caracterização (particular) das hipóteses céticas e de proposições *heavyweight* de modo que, na aplicação de agentes consideradores de hipóteses céticas, os princípios de fecho epistêmicos (E-CLOS 1), (E-CLOS 2), (E-CLOS 3) e (E-CLOS 6) podem ser considerados inválidos ou não-aplicáveis. Esta investigação não pretende solucionar definitivamente o problema do fecho epistêmico numa perspectiva informal, mas sugere fortemente que, seja qual for a solução proposta, ela deverá certamente considerar os avanços da epistemologia formal neste tema.

Referências

ARNER, Douglas. On knowing. **The Philosophical Review**, Vol. 68, n. 1, p. 84-92, jan. 1959.

BARWISE, John; PERRY, John. Situations and attitudes. **Journal of Philosophy**, Vol. 78, n. 11, p. 668-691, nov. 1981.

BELNAP, Alan R.; ANDERSON, Nuel D. **Entailment: the logic of relevance and necessity.** Princeton: Princeton University Press, 1975.

BONJOUR, Lawrence. Nozick, externalism and skepticism. *In*: LUPER-FOY, S. (editor). **The possibility of knowledge.** Lanham: Roman and Littlefield, 1987.

CANTUARIENSIS, Anselmus. Proslogion. The Latin Library. Disponível em http://www.thelatinlibrary.com/anselmproslogion.html. Acesso em: 08/04/2013.

CHELLAS, Brian F. **Modal logic: an introduction.** New York: Cambridge University Press, 1995.

CRUZ, Ângela M. P. **Lógica deôntica paraconsistente: paradoxos e dilemas.** Natal: EDUFRN, 2005.

DE ALMEIDA, Cláudio. Closure, defeasibility and conclusive reasons. **Acta Analytica**, Vol. 22, n. 4, p. 301-319, dez. 2007.

_________. Racionalidade epistêmica e o paradoxo de Moore. **Veritas**, Vol. 54, n. 2, p. 48-73, maio/ago 2009.

DEROSE, Keith. Solving the skeptical problem. **The Philosophical Review**, Vol. 104, n. 1, p. 1-52, jan. 1995.

DOUVEN, Igor. The lottery paradox and our epistemic goal. **Pacific Philosophical Quarterly**, Vol. 89, n. 2, p. 204-225, jun. 2008.

DRETSKE, Fred. I. Epistemic operators. **The Journal of Philosophy**, Vol. 67, n. 24, p. 1007-1023, dez. 1970.

_________. Conclusive Reasons. **Australasian Journal of Philosophy**, Vol. 49,

n. 1, p. 1-22, mai. 1971.

_________. The pragmatic dimension of knowledge. **Philosophical Studies**, Vol. 40, n. 3, p. 363-378, nov. 1981.

_________. Is knowledge closed under known entailment? The case against closure. *In*: STEUP, Mathias (org.). **Contemporary debates in epistemology**. Malden: Blackwell, 2005a, p. 13-26.

_________. Reply to Hawthorne. *In*: STEUP, Mathias (org.). **Contemporary debates in epistemology**. Malden: Blackwell, 2005b, p. 43-46.

_________. Information and closure. **Erkenntnis**, Vol. 4, n. 3, p. 409-413, 2006.

FAGIN, Ronald; HALPERN, Joseph Y. **Belief, Awareness and Limited Reasoning**. *In*: Ninth International Joint Conference on AI. Los Angeles, CA: 1985.

_________. Belief, awareness and limited reasoning. **Artificial Intelligence**. Vol. 34, p. 39-76, 1988.

FAGIN, Ronald *et al*. **Reasoning about knowledge.** Massachusetts: The MIT Press, 2003.

FELDMAN, Richard. Contextualism and skepticism. **Noûs**, Vol. 33, n. 13, p. 91-114, out. 1999.

FITCH, Frederic B. A logic analysis of some value concepts. **The Journal of Symbolic Logic**, Vol. 28, n. 2, p. 135-142, jun. 1963.

FITTING, Melvin; MENDELSOHN, Richard L. **First-order modal logic.** Dordrecht: Kluwer Academic Publishers, 1998.

FUMERTON, Richard, A. Nozick's epistemology. *In*: LUPER-FOY, S. (editor). **The possibility of knowledge**. Lanham: Roman and Littlefield, 1987.

GETTIER, Edmund, L. Is justified true belief knowledge? **Analysis**, Vol. 23, p. 121-123, jun. 1963.

HADLEY, Robert F. **Logical omniscience, semantics, and models of belief**. Burnaby: Simon Fraser University, 1987.

_________. The many uses of "Belief" in AI. **Minds and machines**, Vol. 1, p. 55-73, 1991.

HARMAN, Gilbert; SHERMAN, Brett. Knowledge, assumptions, lotteries. **Philo-**

sophical Issues, Vol. 14, n. 1, p. 492-500, out. 2004.

HAWTHORNE, John. Is knowledge closed under known entailment? The case for closure. *In*: STEUP, Mathias (org.). **Contemporary debates in epistemology.** Malden: Blackwell, 2005, p. 26-43.

HENDRICKS, Vincent; SYMONS, John. Where's the Bridge? Epistemology and Epistemic Logic. **Philosophical Issues**, Vol. 128, n. 1, p. 137-167, mar. 2006.

HENDRICKS, Vincent. **Mainstream and formal epistemology.** New York: Cambridge University Press, 2006.

HINTIKKA, Jaakko. **Knowledge and Belief: an introduction to the logic of the two notions.** New York: Cornell University Press, 1962.

_________. Impossible possible worlds vindicated. **Journal of Philosophical Logic**, Vol. 4, n. 4, p. 475-484. nov. 1975.

HUANG, Zhisheng; KWAST, Karen L. **Awareness, negation and logical omniscience.** *In*: J. Van Eijck (editor): Logic in AI, Proceedings of European Workshop on Logics in Artificial Intelligence. Amsterdam: University of Amsterdam, 1991, p. 282-300.

HUGHES, George E.; CRESSWELL, Maxwell J. **A new introduction to modal logic.** London: Routledge, 1996.

KLEIN, Peter. Closure matters: academic skepticism and easy knowledge. **Philosophical Issues**, Vol. 14, n. 1, p. 165-184, out. 2004.

KONOLIGE, Kurt. **A deduction model of belief and its logics.** Menlo park, 1984. 301 p. Tese. Artificial Intelligence Center, Stanford University.

_________. **What awareness isn't: a sentential view of implicit and explicit belief**. *In*: TARK'86: Proceedings of the 1986 conference on theoretical aspects of reasoning about knowledge. San Francisco: Morgan Kaufmann Publishers Inc., 1986a, p. 241-250.

_________. **A deduction model of belief.** San Francisco: Morgan Kaufmann Publishers Inc., 1986b.

KRIPKE, Saul A. A completeness theorem in modal logic. **The Journal of Symbolic Logic**, Vol. 24, n. 1, p. 1-14, mar. 1959.

KYBURG, Henry L. **Probability and the logic of rational belief.** Middletown:

Wesleyan University Press, 1961.

_________. Conjunctivitis. *In*: Swain, M. (editor): **Induction, acceptance and rational belief.** Dordrecht: Reidel, 1970, p. 55-82.

KVANVIG, Jonathan L. Closure principles. **Philosophy Compass**, Vol. 1, n. 3, p. 256-267, mai. 2006.

LEE, Richard Char-Tung; CHANG, Chin-Liang. **Symbolic logic and mechanical theorem proving.** New York: Academic Press, 1973.

LEVESQUE, Hector J. **A logic of implicit and explicit belief**. *In*: Proceedings of the national conference on artificial intelligence. Austin: AAAI Press, 1984, p. 198-202.

LI, Jinhou. **Is logical omniscience a problem there or not: a critical view.** 30th annual conference of IEEE. China, 2-6/11/2004.

LIPMAN, Barton L. **An axiomatic approach to the logical omniscience problem**. *In*: TARK'94: Proceedings of the 5th conference on theoretical aspects of reasoning about knowledge. San Francisco: Morgan Kaufmann Publishers Inc., 1994. p. 182-196.

LUPER, Steven. Epistemic closure principle. Stanford Encyclopedia of Philosophy, Fall 2012. Disponível em http: //plato.stanford.edu/archives/fall2012/entries/closure-epistemic/. Acesso em 30/01/2013.

MALCOLM, Norman. Knowledge and belief. **Mind** (new series), Vol. 61, n. 242, p. 178-189, abr. 1951.

MCBRIDE, Mark. Is knowledge closed under known entailment? The strange case of Hawthorne's "heavyweight conjuncts" (and other strange cases). **Theoria**, Vol. 75, n. 2, p. 117-128, mai. 2009.

MCINTOSH, J. J. Fitch's factives. **Analysis**, Vol. 44, n. 4, p. 153-158, out. 1984.

MEDEIROS, Stanley K. B. **O problema da onisciência lógica e a lógica das crenças implícitas e explícitas de Levesque**. Comunicação proferida no V Encontro Interinstitucional de Filosofia – CCHLA – UFPB, João Pessoa, agosto de 2007.

_________. **Análise crítica dos impossíveis mundos possíveis de Hintikka.** Comunicação proferida na XVII Semana de Filosofia – CCHLA – UFRN. Natal-RN, outubro de 2007.

_________. **O problema da onisciência lógica: uma discussão sobre soluções e novos problemas.** João Pessoa: 2008, 122 p. Dissertação (Mestrado em Filosofia). Centro de Ciências Humanas, Letras e Artes, Universidade Federal da Paraíba.

MOORE, G. E. **Ethics.** London: Williams & Norgate, 1912.

_________. **A reply to my critics.** *In*: SCHILPP, P. A (editor): The philosophy of G. E. Moore. Evanston: Northwestern Universisty, 1942. p. 541-543.

_________. **Philosophical papers.** London: Allen & Unwin, 1959.

MOORE, Robert C. **Semantical considerations on nonmonotonic logic.** Technical note 284. Menlo Park: SRI International, 1983.

MURPHY, Peter. A strategy for assessing closure. **Erkenntnis**, Vol. 65, n. 3, p. 365-383, nov. 2006.

NOZICK, Robert. **Philosophical explanations.** Cambridge: Harvard Universisty Press, 1981.

OLIN, Doris. A case against closure. **Veritas**, Vol. 50, n. 4, p. 235-247, dez. 2005.

RANTALA, Veikko. Urn models: A new kind of non-standard model for first-order logic. **Journal of Philosophical Logic**, Vol. 4, n. 3, p. 455-474, ago. 1975.

RESCHER, Nicholas. **Epistemic logic: a survey of the logic of knowledge.** Pittsburgh: Universisty of Pittsburgh Press, 2005.

_________. **Epistemology: an introduction to the theory of knowledge.** Albany: State University of New York Press, 2003.

_________. **Unknowability: an inquiry into the limits of knowledge.** Lanham: Lexington Books, 2009.

SHACKEL, Nicholas. Shutting Dretske's door. **Erkenntnis**, Vol. 64, n. 3, p. 393-401, 2006.

STALNAKER, Robert C. **Context and content.** New York: Oxford Universisty Press, 1999.

_________. The problem of logical omniscience, I. **Synthese**, Vol. 89, n. 3, p. 425-440, dez. 1991.

_________. On logics of knowledge and belief. **Philosophical Studies**, Vol 128, n. 1, p. 169-199, mar. 2006.

VARDI, Moshe Y. **On epistemic and logical omniscience.** *In*: TARK'86: Proceedings of the 1986 conference on theoretical aspects of reasoning about knowledge. San Francisco: Morgan Kaufmann Publishers Inc., 1986. p. 293-305.

VON WRIGHT, George H. **Ensayo de logica modal.** Trad. Atílio A. Demarchi. Buenos Aires: Rueda Filosofica, 1970.

WARFIELD, Ted A. When Epistemic Closure Does and Does Not Fail: A Lesson from the History of Epistemology. **Analysis**, Vol. 64, n. 1, p. 35-41, jan. 2004.

WILLIAMS, John N. Moorean absurdity and the intentional structure of assertion. **Analysis**, Vol. 54, n. 3, p. 160-166, jul. 1994.

WILLIAMSON, Timothy. **Knowledge and its Limits.** Oxford: Oxford University Press, 2000.

WHITSEY, Mark. **Logical omniscience: a survey.** Techical report NOTTCS-WP-2003-2. Nottinghan: University of Nottinghan, 2003.