

EPISTEMOLOGIA COMPUTACIONAL: UMA PROVOCAÇÃO

Danilo Fraga Dantas¹

RESUMO

Este artigo discute o uso de simulações de computador em Epistemologia (Epistemologia Computacional). O objetivo do artigo é fundamentar e discutir a ideia de uma Epistemologia Computacional, além de apresentar um exemplo de estudo nesse campo. Na Introdução, discuto as objeções mais comuns aos métodos da Epistemologia Tradicional e à proposta de Quine de uma Epistemologia Naturalizada. Argumento que a Epistemologia Computacional não está sujeita a nenhuma destas objeções. Na Seção 1, apresento uma revisão bibliográfica dos estudos em Epistemologia Computacional (tanto em Epistemologia individualista quanto em Epistemologia Social) e discuto a estrutura geral destes estudos. Na Seção 2, apresento alguns resultados de um estudo em Epistemologia Computacional que realizei em minha tese de doutorado.

Palavras-chave: Epistemologia Computacional. Epistemologia Naturalizada. Epistemologia Social. Epistemologia Bayesiana. Racionalidade epistêmica. Lógica Não-Monotônica.

ABSTRACT

This paper is about the use of computer simulations in Epistemology (Computational Epistemology). The goal of the paper is to ground and discuss the idea of a Computational Epistemology, and to present an example of study in this field. In the Introduction, I discuss the most common objections to the methods of Traditional Epistemology and to Quine's Naturalized Epistemology. I argue that Computational Epistemology is not subject to any of these objections. In Section 1, I review the literature on Computational Epistemology (both in individualistic Epistemology and in Social Epistemology) and discuss the general structure of these studies. In Section 2, I present some results of my PhD dissertation on Computational Epistemology.

Keywords: Computational Epistemology. Naturalized Epistemology. Social Epistemology. Bayesian Epistemology. Epistemic rationality. Nonmonotonic Logic.

¹ Estagiário pós-doutoral do PPG-F na UFSM. E-mail: dfdantas@ucdavis.edu.

Introdução

Nas últimas décadas, a natureza da Epistemologia tem sido muito discutida. Entre outras questões, tem sido discutido se a Epistemologia é uma disciplina: (i) predominantemente a priori ou se deveria utilizar métodos e resultados científicos; (ii) normativa ou completamente descritiva; (iii) autônoma ou parte integrante de outra disciplina. Essas questões foram trazidas para o centro da discussão pelo artigo ‘Epistemology Naturalized’. Neste artigo, Quine recomenda que os epistemólogos utilizem métodos e resultados científicos para descrever o modo como humanos, de fato, formam crenças:

A relação entre a entrada diminuta [ex. de dados dos sentidos] e a saída torrencial [ex. de crenças sobre o mundo externo] é a relação que nos sentimos estimulados a estudar mais ou menos pelas mesmas razões que sempre serviram de estímulo à Epistemologia... Mas uma diferença sensível entre a antiga epistemologia e o novo programa epistemológico está em que agora temos todo o direito de recorrer livremente à Psicologia empírica (QUINE, 1969, p. 288, *adendos meus*).

Nesse artigo, Quine também defende a ‘tese da substituição’: “A Epistemologia, ou algo que se lhe assemelhe, encontra o seu lugar simplesmente como um capítulo da Psicologia e, portanto, da ciência natural” (QUINE, 1969, p. 287). A Epistemologia Naturalizada (EN) seria uma disciplina empírica e descritiva, que não passaria de um “capítulo da Psicologia” (não-autônoma). A Epistemologia Tradicional (ET), por outro lado, é tida como uma disciplina predominantemente a priori², normativa e autônoma (cf. RYSIEW, 2020).

A discussão acerca da EN foi intensificada nas últimas décadas, quando os métodos da ET foram colocados em questão. Um dos projetos centrais da ET é o de oferecer uma análise da noção de conhecimento, o que é tipicamente entendido como um conjunto de condições necessárias e suficientes para a aplicação dessa noção. Na execução deste projeto, epistemólogos comumente utilizam o ‘método de casos’: eles descrevem um ‘caso’ de um sujeito numa determinada situação epistêmica, consultam suas intui-

² A ET é geralmente identificada com a imagem de um filósofo avançando teses sentado numa poltrona. Não necessariamente tudo o que se faz da poltrona é a priori. Epistemólogos tradicionais, por vezes, também usam conhecimento de senso comum (ex. linguístico) ou mesmo alguns resultados (mas não métodos) científicos.

ções sobre se esse é um caso em que o sujeito tem conhecimento e decidem se este é um contraexemplo a uma proposta de análise. Os exemplos paradigmáticos desse método são os contraexemplos do estilo-Gettier para a análise clássica do conhecimento (cf. BLOWN et al., 2018).

O método de casos tem dois problemas sérios. O primeiro problema, que diz respeito ao uso evidencial de intuições epistêmicas, foi evidenciado pelos resultados da Filosofia Experimental (uma abordagem que procura lidar com questões filosóficas a partir de métodos associados às Ciências Sociais). Os resultados Filosofia Experimental sugerem que intuições epistêmicas são influenciadas por fatores demográficos, como o gênero e a etnicidade (cf. KNOBE; NICHOLS, 2017, para uma revisão). Um dos trabalhos mais importantes dessa vertente é o de Weinberg et al. (2001)³. Se intuições são influenciadas por fatores demográficos, seu uso evidencial poderia ter como consequência com que os resultados em ET estejam enviesados pela demografia muito homogênea dos epistemólogos (homens, brancos, de meia idade, de classe média/alta, etc.).

Na literatura, existem tentativas de defender o uso evidencial de intuições epistêmicas em ET. Uma maneira de responder aos resultados da Filosofia Experimental é afirmar que sujeitos de diferentes populações podem estar interpretando diferentemente os cenários propostos. Nesse caso, poderíamos explicar as respostas diferentes argumentando que esses sujeitos estão respondendo a questões diferentes (cf. SOSA, 2009). Outra maneira de responder a esses resultados é questionar os próprios resultados. Por exemplo, apesar de Weinberg et al. (2001) terem encontrado uma variação significativa entre as intuições de sujeitos de culturas diferentes, estudos posteriores (ex. SEYEDSAYAMDOST, 2015, e KIM; YUAN, 2015) falharam em replicar esses achados. Não quero entrar nessa questão. Quero apenas apontar que o uso evidencial de intuições epistêmicas em ET é controverso.

O segundo problema do método de casos diz respeito a seu foco em contraexemplos. Esse foco acontece porque uma análise é tomada como um

³ Em um de seus experimentos mais célebres, Weinberg et al. apresentaram a 89 sujeitos, divididos entre ‘ocidentais’ e ‘leste-asiáticos’, os contraexemplos de Gettier. Enquanto apenas ~26% dos ocidentais classificaram esses como casos de conhecimento, ~57% (ou seja, a maioria) dos leste-asiáticos o fizeram.

conjunto de condições necessárias e suficientes para a aplicação de um conceito. Nesse caso, um contraexemplo seria suficiente para invalidar uma análise. O foco em contraexemplos tem consequências desafortunadas. No intuito de evitar os contraexemplos anteriores, análises tendem a ficar mais complexas, incorrendo no uso de pseudoformalismos (ex. na condição de segurança, WILLIAMSON, 2000, p. 147)⁴. Contraexemplos, por sua vez, tendem a ficar mais complexos e a descrever situações menos cotidianas (ex. COMESAÑA, 2005)⁵. Porém, nossas intuições tendem a ser menos confiáveis exatamente quando lidamos com situações menos cotidianas. Um problema ainda mais sério deve-se à ideia de que um contraexemplo é capaz de invalidar uma hipótese (de análise). Nesse caso, hipóteses que funcionam bem para a maioria dos casos podem ser descartadas com base num único contraexemplo, mesmo que artificial.

Nesse contexto, parece ser uma boa ideia utilizar métodos científicos (e bem estabelecidos) em Epistemologia. Mas há um problema: a Epistemologia tem um caráter normativo que esses métodos não parecem sustentar. Por ‘caráter normativo’, quero dizer que ao menos algumas das afirmações em Epistemologia são avaliativas. Por exemplo, dizer que uma crença é justificada é dizer que esta crença é permissível de um ponto de vista epistêmico. Alguns defensores da ET mantêm que a Epistemologia também seria normativa por ser prescritiva, no sentido de nos dizer como nós deveríamos formar nossas crenças, etc. A Psicologia Cognitiva estuda o modo como humanos, de fato, formam crenças. É bem sabido que o modo como humanos formam crenças exhibe vieses⁶ (KAHNEMAN, 2012), entre outros vícios epistêmicos. Nesse sentido, é difícil conceber como dados desse tipo poderi-

⁴ A condição de segurança é geralmente descrita da seguinte maneira: Se *S* tem conhecimento, então *S* não poderia estar facilmente errado numa situação *similar*. Apesar da aparente clareza, esta cláusula esconde sua vagueza sob noção de similaridade. Quão similar é suficientemente similar? Similaridade sob qual aspecto?

⁵ Comesaña (2005) apresenta o seguinte caso como sendo um contraexemplo à condição de segurança: Judy foi contratada para orientar os convidados sobre a localização de uma festa. Ela deve dar instruções corretas aos convidados, mas, se Michael aparecer, ela deve lhe dizer (falsamente) que a festa foi transferida para outro local (a ideia é que Michael não apareça na festa). Suponha que que um dos convidados, apesar de não ter decidido, esteve muito próximo de decidir usar uma fantasia muito realista de Michael para ir à festa. Nesse caso, sua crença sobre a localização da festa (que é baseada unicamente no testemunho de Judy) seria conhecimento, apesar de poder facilmente ser falsa. Mas será esta uma situação é suficientemente similar?

⁶ Vieses são desvios sistemáticos dos cânones da lógica clássica ou de outro modelo de raciocínio correto.

am, por si só, sustentar uma abordagem normativa (avaliativa ou prescritiva). Esse tipo de argumentação é conhecida como “a objeção normativa” (cf. KIM, 1988, p. 389)⁷.

Em resposta à objeção normativa, Quine propõe que a normatividade da EN deveria ser entendida a partir de uma analogia com a engenharia.

A naturalização da Epistemologia não descarta o normativo e adota a descrição indiscriminada de processos ocorrentes. Para mim, a Epistemologia normativa é um ramo da engenharia. É a tecnologia da busca pela verdade ou, utilizando termos epistêmicos mais cautelosos, da predição... Não há aqui a questão sobre valor fundamental, como na moral; trata-se de eficiência em relação a uma finalidade oculta, verdade ou predição. O normativo aqui, como em qualquer engenharia, torna-se descritivo quando o parâmetro final é expresso (QUINE, 1998, pp. 664-665, *tradução minha*).

A ideia parece ser que essa analogia explique como se daria a normatividade de EN⁸. Num certo sentido, a engenharia é uma disciplina normativa. Por exemplo, suponha que queiramos construir uma ponte. Nesse caso, poderíamos contratar engenheiros de pontes que empregariam métodos e resultados da física, geologia, etc., para distinguir entre bons e maus projetos de construção (avaliação) para, então, prescrever-nos o melhor projeto (exemplo de WRENN, 2006, p. 68). Porém, há uma limitação nessa estratégia, apontada por Wrenn (2006): a normatividade em Epistemologia seria *categórica*, independente de objetivos individuais (ex. construir uma ponte), mas a analogia seria capaz apenas de fundamentar uma normatividade *hipotética* (dependente de objetivos individuais).

Para ser justo, nem todos os epistemólogos naturalistas concordam que a Epistemologia deva lidar com normatividade categórica. Por exemplo, Stich (1993, p. 9), argumenta que não há objetivos epistêmicos “especiais” e que, conseqüentemente, os valores relevantes são aqueles que promovem

⁷ Para Kim, Quine propõe ignorar a relação de suporte epistêmico e investigar as conexões causais entre nossa evidência sensória e nossas crenças sobre o mundo (essas seriam relações diferentes dos mesmos *relata*).

⁸ De fato, existe uma tensão na posição de Quine. Por um lado, ele afirma que “não há aqui a questão sobre valor fundamental” e que “o normativo aqui, como em qualquer engenharia, torna-se descritivo *quando o parâmetro final é expresso*”, o que suscita não haver um valor epistêmico fundamental. Por outro lado, ele fala em “uma finalidade oculta, verdade ou predição”, o que suscita que a verdade ou predição é, no final das contas, o tal “valor fundamental” ou “parâmetro final”. Como não está claro que a analogia com a engenharia sustentaria que esta seria a “finalidade oculta” da atividade epistêmica. Por isso, atenho-me à primeira opção.

objetivos individuais (ex. a felicidade). Porém, esta posição é instável. O alcance efetivo de um objetivo individual qualquer depende da posse (pelo indivíduo) de uma capacidade cognitiva capaz de avaliar confiavelmente os melhores meios de alcançar este objetivo. Mas, dado que capacidades cognitivas confiáveis são necessárias para a seleção dos melhores meios para alcançar um objetivo individual, a avaliação dessas capacidades (ex. se elas, de fato, selecionam os melhores meios) deve ser independente de qualquer objetivo individual específico (KORNBLITH, 2002, p. 158).

A posição mais controversa de Quine (1969) é certamente a tese da substituição. Essa tese é majoritariamente rejeitada tanto por epistemólogos tradicionais, quanto por epistemólogos com tendências mais naturalistas (cf. FOLEY, 1994). Minha crítica à tese da substituição é que, mesmo deixando de lado seus aspectos normativos, a Epistemologia parece ter um domínio de investigação modalmente mais amplo do que, por exemplo, a Psicologia Cognitiva. A Epistemologia está interessada (ex. em avaliar) não apenas as práticas epistêmicas atualmente utilizadas, mas também em (avaliar ou prescrever) práticas possíveis (ex. para humanos), mas nunca utilizadas. Por isso, a redução da Epistemologia a, por exemplo, à Psicologia Cognitiva, em vez de resolver os problemas da Epistemologia, apenas diminui (artificialmente) seu escopo.

Em suma, os epistemólogos tradicionais estão corretos em apontar para a normatividade e a autonomia da Epistemologia. Porém, o uso evidencial de intuições e o foco em contraexemplos na ET realmente são problemáticos. Quine, por outro lado, está correto ao apontar para a necessidade de se utilizar métodos e resultados científicos (e bem estabelecidos) em Epistemologia. Porém, a EN falha em preservar a normatividade e a autonomia da Epistemologia. No que se segue, apresento a ideia de uma Epistemologia Computacional (EC), que evita ambos os problemas.

Rumo à EC

EC é uma abordagem procedural, no sentido de Pollock (1999). Num estudo típico de EC, é construída uma simulação computacional em que agentes artificiais (construídos para exibir uma prática epistêmica de in-

teresse) atualizam suas crenças em interação com o ambiente (ex. outros agentes) gerados aleatoriamente a partir de alguns parâmetros fixados. O objetivo é estudar práticas epistêmicas a partir de seu impacto nas crenças pelos agentes. Esta abordagem é adequada ao estudo da noção de racionalidade (epistêmica), entendida aqui como a propriedade de uma prática de maximizar algum tipo de valor epistêmico. O mesmo não pode ser dito em relação ao estudo da análise de ‘conhecimento’. Existe uma noção de “justificação procedural” (POLLOCK, 1999, p. 386): uma crença é justificada se e somente foi adquirida ou é mantida com base em procedimentos epistêmicos racionais. Porém, esta não é a noção de justificação relevante para a análise de ‘conhecimento’⁹.

Racionalidade está relacionada com a maximização de algum tipo de valor epistêmico. Na literatura, diferentes características (ex. de um conjunto de crenças) são tomadas como fontes de valor epistêmico: consistência ou coerência, fechamento sob alguma operação (ex. consequência lógica), quantidade/qualidade do suporte evidencial, etc. Nas últimas décadas, porém, epistemólogos têm argumentado que a fonte fundamental de valor epistêmico de uma prática epistêmica está relacionada a quanto ela maximiza as crenças verdadeiras e minimiza as falsas (cf. TITELBAUM, no prelo, cap. 10). Esta posição é chamada de ‘veritismo’. Não necessariamente toda EC é veritista, porém é um fato que a maioria dos estudos publicados segue essa tendência (cf. Seção 1). O apelo do uso de métodos computacionais numa abordagem veritista vem do fato de, numa simulação, o pesquisador tem controle não apenas das práticas epistêmicas, mas também do ambiente, de modo a sempre ser possível determinar o valor de verdade de crenças dos sujeitos.

É geralmente aceito que a racionalidade tem um caráter subjetivo, que estaria relacionado à probabilidade (subjetiva ou evidencial) de crenças serem verdadeiras. Por esse motivo, a racionalidade de uma prática epistêmica é geralmente associada a seu valor epistêmico *esperado*, em vez de seu valor absoluto (ex. na Teoria da Utilidade Epistêmica, TUE). Nesse contexto, seria uma estratégia viável em Epistemologia avaliar e prescrever práti-

⁹ Os contraexemplos de Gettier podem ser entendidos como casos em que alguém tem uma crença verdadeira formada de maneira racional, mas falha em ter conhecimento por motivos que escapam a seu controle.

cas epistêmicas a partir de seu valor epistêmico *esperado*. No caso de práticas epistêmicas muito simples (ex. adoção de uma ou de poucas crenças) e de ambientes muito simples e abstratos (ex. compostos de uma ou de poucas proposições), sua análise e prescrição podem ser feitas ‘analiticamente’ (através de argumentos a priori). Por exemplo, a TUE tem produzido argumentos em favor do probabilismo: graus de crença que expressam funções de probabilidade maximizam o valor epistêmico esperado em relação aos que não expressam (TITELBAUM, no prelo, cap. 10).

Esta estratégia não parece estar disponível para práticas epistêmicas e situações complexas. Por exemplo, qual dessas práticas é epistemicamente mais deletéria para uma comunidade: a mentira ou a “falação de merda” (cf. FRANKFURT, 2005)¹⁰? Não creio que exista um argumento conclusivo para qualquer das conclusões. Em primeiro lugar, é difícil medir o valor epistêmico absoluto da mentira e da falação de merda numa comunidade atual¹¹. O valor epistêmico *esperado* dessas práticas é ainda mais difícil de medir, pois esse valor depende não apenas de seu impacto numa situação atual, mas também de seu impacto em situações não-atuais. Em casos como esse, parece ser uma boa ideia construir uma simulação em que agentes (que mentem ou falam merda) atualizam suas crenças interagindo com os outros agentes em redes de agentes (ambientes) geradas aleatoriamente. Se o número de ambiente for alto, a média geral de valor epistêmico deve se aproximar de seu valor esperado. Em resumo, EC é uma alternativa viável para o estudo da racionalidade de práticas e situações mais complexas (ex. práticas epistêmicas e situações mais realistas).

Intuições epistêmicas têm um uso em EC, mas este uso não é evidencial, é heurístico. Num estudo em EC, simulamos práticas epistêmicas hipoteticamente racionais. Essas hipóteses podem ter diferentes origens, incluindo nossas intuições epistêmicas (outras possibilidades são discutidas na Seção 1). Porém, seja qual for sua origem, hipóteses são testadas partir de

¹⁰ “Falar merda”, aqui, é um termo técnico introduzido por Frankfurt (2005). Diferentemente da mentira, que envolve afirmar algo que se acredita ser falso, falar merda se caracteriza pela falta de interesse na verdade do que se fala. Frankfurt hipotetiza que a falação de merda é mais epistemicamente danosa que a mentira.

¹¹ Se não por outros motivos, quando da afirmação de uma falsidade (ex. compartilhamento de fake news), é difícil distinguir entre as possibilidades de mentira, falação de merda e engano genuíno do sujeito que afirma.

simulações de computador. O uso heurístico de intuições é justificado porque um agente de uma simulação pode ser visto com uma versão de nós mesmos e nossas intuições estariam relacionadas a quais práticas adotaríamos se estivéssemos naquela situação (supostamente, temos alguma competência sobre como raciocinar). Porém, hipóteses geradas a partir de nossas intuições não são garantidas de estarem corretas porque nossas intuições são menos confiáveis em situações complexas¹².

A EC não se fia em contraexemplos para avaliar hipóteses. O fato da EC lidar com a *média* do valor epistêmico numa variedade de situações geradas aleatoriamente tem como consequência que um contraexemplo não seja suficiente para a rejeição de uma hipótese. Por exemplo, uma hipótese com alto valor epistêmico num grande número de situações não seria rejeitada devido a um valor epistêmico baixo numa situação específica. Além disso, a necessidade de implementar hipóteses em simulações de computador exige que estas sejam definidas rigorosamente, dificultando o uso de pseudoformalismos. Por fim, os métodos utilizados em EC são bem estabelecidos. Métodos similares são utilizados, de maneira bem-sucedida, em outras disciplinas (cf. WINSBERG, 2019 e GRIM; SINGER, 2020).

Diferentemente da EN, a EC é (parte de) um projeto de pesquisa normativo. A pesquisa em EC parte do pressuposto de que há um valor epistêmico categórico a ser maximizado. O estabelecimento desse valor, porém, é resistente a argumentos naturalistas. Por exemplo, a argumentação mais comum em favor da verdade ser um valor categórico é derivada da ideia de que é constitutivo da noção de crença que crenças têm como objetivo a verdade (SHAH, 2003). Nesse caso, a EC pode ser entendida como uma “Epistemologia enquanto engenharia”, mas com o “parâmetro final” fixado por vias tradicionais. Há tentativas naturalistas de fixar esse parâmetro. Por exemplo, como exposto acima, Kornblith (2002) argumenta que, dado que capacidades cognitivas são recrutadas para avaliar meios para alcançar objetivos, a verdade seria um valor para todos que tenham objetivos (ex. huma-

¹² Cummins (1998) aponta que um dos problemas do uso evidencial de intuições epistêmicas é a falta de um método confiável de calibragem para estas intuições. A EC pode fazer o papel desse método. Por exemplo, se temos a intuição de que o princípio da explosão não descreve uma prática epistêmica racional, podemos comparar um agente cujo raciocínio exiba esse princípio com outro que exiba um princípio paraconsistente.

nos). Mas não é óbvio que a norma ‘se você tem objetivos, deve valorizar a verdade’ seja categórica. Em qualquer um dos casos, a EC é (parte de) um projeto normativo.

Por fim, a EC não ameaça à autonomia da Epistemologia. É bem verdade que simulações são utilizadas em disciplinas como a Psicologia Cognitiva e a Inteligência Artificial (IA), mas isso não faz da EC parte dessas disciplinas. EC e Psicologia Cognitiva utilizam métodos computacionais para estudar o raciocínio, mas com objetivos diferentes: enquanto o objetivo da Psicologia Cognitiva é descrever o modo como humanos, de fato, raciocinam o objetivo da EC é avaliar e prescrever práticas epistêmicas. A EC e a IA parecem estar ambas interessadas em estudar padrões de inferência ótimos. Mas a IA está interessada em encontrar estratégias ótimas para solucionar *problemas específicos*. A EC, por outro lado, está interessada em avaliar e prescrever mecanismos mais ou menos gerais de raciocínio. Em resumo, a EC pode ser entendida como uma forma moderada de EN, que usa métodos e resultados científicos, mas que preserva a normatividade e autonomia da Epistemologia.

1. Epistemologia computacional

Historicamente, a Epistemologia tem se concentrado no caso individualista (em que sujeitos são estudados de maneira mais ou menos isolada de outros sujeitos). Porém, uma Epistemologia que trate apenas de sujeitos isolados é, no mínimo, incompleta, já que parte considerável do conhecimento de que, de fato, possuímos resulta da interação com outros sujeitos. Esse tipo de consideração levou ao desenvolvimento da Epistemologia Social (cf. GOLDMAN, 1999, para uma introdução). Ironicamente, a EC social está num estágio mais avançado que a EC individualista. Um dos motivos para isto é o fato de os modelos da EC social terem um grau muito maior de abstração que os da EC individualista, o que implica em maior facilidade de implementação. Por isso, há plataformas de interface amigável para estudos em EC social (ex. Laputa), mas há nada parecido para a EC individualista.

EC individualista

John Pollock é um dos precursores do uso ferramentas computacionais em Epistemologia individualista. Pollock implementou suas teorias em sobre raciocínio anulável, teoria da decisão, etc. num agente de IA chamado OSCAR (johnpollock.us/ftp/OSCAR-web-page). Além disso, Pollock discutiu a implementação de teorias epistemológicas em agentes de IA em diversos artigos (ex. POLLOCK, 1999) e nos livros ‘How to build a person’ (1989) e ‘Cognitive carpentry’ (1995). Apesar de sua importância para o campo, creio que Pollock não se dedicou à EC no sentido discutido na Introdução. O ponto é que, em geral, Pollock utilizou OSCAR como uma maneira de expressar claramente teorias (ex. evitando pseudoformalismos), mas não costumava testar suas hipóteses a partir da simulação de hipóteses em problemas gerados aleatoriamente.

Comumente, humanos (ex. cientistas) precisam decidir entre hipóteses com base na evidência disponível. Uma hipótese de como essa decisão pode ser feita racionalmente é que julgamos qual hipótese melhor explica a evidência disponível para, então, fazermos uma inferência para a melhor explicação (IME). A ideia de IME enfrenta dois desafios. Primeiramente, como a IME deve ser entendida e formalizada? Além disso, quão bem a IME maximiza o valor epistêmico (ex. em comparação a outras formas de raciocínio sob incerteza)? Glass (2012) investiga diferentes formalizações da IME incluindo uma baseada na medida de coerência por sobreposição (GLASS, 2002; OLSSON, 2002). Ele construiu simulações em que agentes devem decidir qual a melhor explicação para a evidência disponível entre um conjunto de hipóteses mutuamente exclusivas e exaustivas (cada qual com diferentes probabilidade inicial e verossimilhança em relação à evidência). O objetivo era medir o quanto cada agente seleciona a hipótese correta. Os resultados sugerem que a escolha baseada na medida de coerência por sobreposição maximiza o valor epistêmico em comparação com as rivais¹³.

¹³ Angere (2007) também usa métodos computacionais para comparar diferentes medidas de coerência e conclui que a medida de coerência por sobreposição maximiza o valor epistêmico relativamente às rivais.

Em casos em que a evidência é incerta, porém, o resultado depende do grau de incerteza adotado (cf. GLASS, 2012, p. 422)¹⁴.

Variedades não Bayesianas de IME são tomadas como formas irracionais de atualização de crenças na medida em que diferem na condicionalização Bayesiana. Nas últimas décadas, o principal argumento para essa conclusão parte do fato de que a condicionalização minimiza a inacurácia esperada (ou seja, maximiza o valor epistêmico esperado) em relação a todos os outros métodos de atualização de crenças (cf. TITELBAUM, no prelo, cap. 10). Douven (2013) constrói simulações em que agentes, um Bayesiano e um explicacionista, assistem a uma mesma sequência de lançamentos de uma moeda e devem descobrir seu grau de enviesamento. Os resultados sugerem que, além de convergir mais rapidamente, o agente explicacionista minimiza a inacurácia *na maioria dos casos*. Porém, quando a inacurácia do explicacionista é maior, ela é muito maior (a condicionalização minimiza a inacurácia *esperada*). Douven conclui que, apesar da condicionalização minimizar a inacurácia esperada, a IME minimiza a inacurácia num sentido relevante (na maioria dos casos)¹⁵.

Outro exemplo de estudo em EC individualista é Dantas (2016, cap. 5), em que um agente Bayesiano e um agente que raciocina não-monotonicamente lidam com raciocínio incerto num ambiente moderadamente mais realista que a média da EC (cf. Seção 2).

EC social

Um dos modelos computacionais mais populares em EC social é o Hegselmann–Krause (HK), que foi apresentado por Hegselmann e Krause (2002) e desenvolvido em (2006). Inicialmente, HK foi aplicado a questões em Ciências Sociais (ex. sob quais circunstâncias uma comunidade converge para uma opinião comum?), mas depois o modelo foi aplicado em EC so-

¹⁴ Glass (2018) propõe uma nova medida de coerência (coerência por produto) e dá prosseguimento ao estudo anterior incluindo essa medida. Nesse estudo, a coerência por produto se sai melhor que a por sobreposição.

¹⁵ Trpin e Pellert (2018) testam casos em que agentes têm evidência incerta sobre os resultados dos lançamentos. Ele compara um agente que utiliza condicionalização de Jeffrey (versão da condicionalização que lida com evidência incerta) e um agente que utiliza uma variação da IME. Os resultados sugerem que o agente explicacionista, além de convergir mais rapidamente para uma resposta, minimiza a inacurácia na maioria dos casos e lida melhor com maior incerteza da evidência.

cial (ex. convergência para a verdade). HK representa uma comunidade de agentes que tem como objetivo determinar o valor de uma variável no intervalo $[0, 1]$. As opiniões iniciais dos agentes são geradas aleatoriamente (com distribuição uniforme) nesse intervalo. Os agentes atualizam suas crenças por deliberação, calculando a média entre a sua opinião atual e a dos outros agentes que tenham opiniões suficiente próximas à sua. Eles ‘dividem a diferença’. Os agentes também fazem ‘atualização evidencial’, a partir dos resultados de investigação individuais acerca do valor da variável. Uma revisão mais detalhada sobre o uso de HK em EC social está em Douven (2020).

Usando uma versão modificada de HK, Douven (2010) argumenta que não existe uma resposta geral para a questão sobre se uma comunidade (ex. a científica) deve adotar a prática de ‘dividir a diferença’ em casos de desacordo entre pares. Os resultados sugerem que comunidades que ‘dividem a diferença’ tendem a convergir mais rapidamente a uma distância moderada da verdade. Por outro lado, comunidades que não ‘dividem a diferença’ (que só executam atualização evidencial) apesar de convergirem mais lentamente, convergem para uma distância menor da verdade. Crosscombe e Lawry (2016) usam HK para investigar convergência em crenças vagas. Os resultados das simulações sugerem que comunidades que ‘dividem a diferença’ convergem para crenças menos vagas do que aquelas que só executam atualização evidencial.

Eric Olsson e Sofi Angere desenvolveram Laputa (luiq.lu.se/portfolio-item/laputa), um modelo para investigações em EC social e uma plataforma com interface amigável para gerar simulações neste modelo. Olsson (2011) apresenta as principais características de Laputa: agentes são modelados como nós num grafo, em que arestas modelam canais de comunicação. Tal qual em HK, agentes atualizam crenças através de deliberação e investigação. Porém, em Laputa, é possível modificar diversos parâmetros que são fixos em HK. Por exemplo, é possível escolher a estrutura da rede (quais agentes se comunicam com quais agentes). Mas também é possível fixar outros parâmetros como: o limiar de certeza para que cada agente afirme sua opinião através dos canais de comunicação, o grau de confiança que cada agente tem em relação a cada outro agente, a chance de agentes fazerem in-

investigações individuais (e a confiabilidade dessas investigações), etc. Um ponto importante do modelo é que, a cada interação, além de atualizar suas próprias crenças, os agentes atualizam o grau de confiança nos outros agentes (ambos por condicionalização).

Hahn et al (2018) investigam como diferentes padrões de conectividade em redes sociais afetam na probabilidade do grupo convergir para a verdade (competência coletiva). Hahn et al utilizam Laputa para simular 36 redes sociais, com 10, 15 ou 18 agentes, e diferentes padrões de conectividade. Os resultados das simulações sugerem que quanto a maior conectividade da rede: (i) mais rapidamente ela converge para um estado estável, porém, contrariamente ao senso comum, (ii) menor é sua competência coletiva¹⁶. No caso de redes com o mesmo grau de conectividade, a diferença entre competências coletivas é explicada pelo coeficiente de agrupamento (quanto maior, pior a competência coletiva). Um coeficiente de agrupamento maior está relacionado a grupos com muita conexão interna, mas pouca com outros grupos (*clustering*). Redes com coeficientes de agrupamento alto exibem fenômenos como polarização e câmaras de eco¹⁷. Pallavicini et al. (2018) usam Laputa para estudar condições gerais em que polarização ocorre. Baumgaertner (2014) usa Netlogo (ccl.northwestern.edu/netlogo), outra plataforma amigável para simulações em EC social, para estudar o papel da imparcialidade em evitar a ocorrência de câmaras de eco.

Discussão

De maneira geral, um estudo em EC é composto por quatro estágios. O primeiro estágio é a escolha e a descrição de uma classe de problemas. O segundo estágio é a escolha das hipóteses de práticas epistêmicas (que exibiriam racionalidade em relação a problemas da classe de interesse) a serem testadas. O terceiro estágio é a implementação de simulações de computador, em que agentes artificiais, construídos para exibir as práticas epistêmicas de interesse, lidam com problemas na classe de interesse. O último estágio diz respeito à coleta e à análise dos dados extraídos das simulações.

¹⁶ Zollman (2007) chega a resultados semelhantes utilizando outro modelo.

¹⁷ Câmaras de eco são situações em que agentes têm suas crenças ‘ecoadas’ de volta por outros agentes dando a impressão geral de que eles estão corretos.

Problema

A medida de valor epistêmico médio relevante para considerações sobre racionalidade deveria (idealmente) levar em conta *todas* as situações possíveis. Mas uma investigação computacional que leve em consideração todas (infinitas) as situações possíveis é inviável. Por outro lado, um estudo consequente sobre racionalidade não pode estar ligado à análise de uma ou de poucas situações. Um meio termo entre esses dois extremos é fazer com que um estudo em EC leve em conta uma classe de problemas¹⁸. A escolha dessa classe segue duas diretrizes: (i) os problemas na classe devem ser o mais abstratos possível, mas, ainda assim, (ii) exibir características comuns a uma grande quantidade de problemas reais.

O preceito (i) tem como objetivo de viabilizar a implementação dos problemas desta classe em simulações. Mesmo assim, não é viável simular todos os problemas numa classe (que pode ser infinita). Para que o estudo não fique ligado aos problemas específicos que foram, de fato, simulados, mas à classe, é de praxe gerar (aleatoriamente) um grande número de problemas nesta classe (geralmente, milhares) e tirar a média dos resultados das simulações. A diretriz (ii) diz respeito à relação entre resultados sobre uma classe de problemas e a noção geral de racionalidade. Resultados sobre uma classe de problemas são relevantes para a noção de racionalidade somente na medida em que os problemas desta classe exibam características comuns a uma grande quantidade de problemas reais. Parte importante do mérito de um estudo em EC depende da escolha da classe de problemas¹⁹.

Uma classe de problemas determina um ou mais objetivos. O propósito da EC é estudar a noção de racionalidade, de modo que um dos objetivos deve sempre ser a maximização do valor epistêmico. O objetivo

¹⁸ Um pouco de vocabulário: um ambiente é composto por um conjunto de estados possíveis (incluindo os estados iniciais possíveis dos agentes), um conjunto das ações possíveis e uma função que retorna o estado resultante de cada ação. Um problema é composto por um ambiente, os estados iniciais de cada agente e um ou mais objetivos. Uma classe de problemas é composta pelos problemas que diferem apenas nos estados iniciais dos agentes. A proposta de formalização (que sigo parcialmente) é Russell and Norvig (2010, p. 66).

¹⁹ Por exemplo, creio que a relevância dos resultados de Hahn et al (2018) é fica limitada pela escolha de uma classe de problemas em que as crenças iniciais dos agentes são, em média, tão acuradas quanto um processo aleatório e a confiança entre agentes é inicialmente total. Hahn et al argumentam que essas características são exibidas na comunicação online entre agentes anônimos, em que não há informação sobre a confiabilidade dos outros e em que todos são ignorantes sobre a questão (p. 1518). Essas características não são gerais.

epistêmico é expresso de diferentes formas: decidir qual hipótese é verdadeira (GLASS, 2012), quão enviesada é uma moeda (DOUVEN, 2013), o valor de uma variável (HK), etc. Porém, a classe de problemas pode também definir objetivos não-epistêmicos. Supostamente, uma prática epistêmica racional é interruptível (POLLOCK, 1995, p. 272), no sentido de que, se um agente (ou grupo de) precisasse interromper a execução da prática e agir num dado momento, seria razoável agir com base nas crenças aceitas até o momento. O uso de objetivos não-epistêmicos pode ajudar a lidar com questões de interruptibilidade. Objetivos não-epistêmicos geralmente não são expressos claramente em estudos de EC, mas argumentações acerca de ‘convergência rápida’ estão ligadas à interruptibilidade²⁰. Dantas (2016, cap. 5) investiga uma classe de problemas com objetivos não-epistêmicos expressos claramente (cf. Seção 2).

Hipóteses

A hipótese de uma prática epistêmica racional a ser investigada pode ter diferentes origens. Podemos (i) usar nossas intuições sobre racionalidade, (ii) testar teorias preexistentes sobre racionalidade (ex. na Epistemologia) ou (iii) partir de resultados científicos sobre práticas epistêmicas em humanos ou outros animais. Opção (i) foi discutida na Introdução.

A opção (ii) é a mais popular em EC. Por exemplo, Glass (2012) parte de formalizações da IME existentes na literatura (ex. OLSSON, 2002). Acredito que esse seja uma boa prática. Geralmente, as hipóteses disponíveis na literatura de Epistemologia já foram refinadas por anos de discussão. Além disso, esse tipo de prática é interessante porque um estudo em EC pode ajudar a resolver disputas antigas em Epistemologia (ou, ao menos, enriquecer essas discussões). Os resultados de Douven (2013) e Trpin e Pellert (2018), por exemplo, são um contraponto interessante aos resultados da

²⁰ Esse parece ser o ponto de Douven (2013) quando ele se questiona se devemos ter o objetivo de minimizar a inacurácia a longo prazo, “quão distante esse futuro possa ser”, ou se é melhor ter crenças moderadamente acuradas num prazo menor. Ele continua: “É certamente imaginável que, por razões puramente epistêmicas, alguém opte pela segunda opção, supondo que há uma escolha a ser feita. Afinal, somos todos *curiosos* sobre a verdade e a curiosidade tipicamente vem com um senso de urgência” (p. 429, *tradução minha*). Douven faz algumas simulações usando uma medida que mede “acurácia gradativa” para lidar com a interruptibilidade (DOUVEN, 2013, p. 430). Não creio, porém, que o “o senso de urgência da curiosidade” seja um sentimento propriamente epistêmico. Por isso, acho mais honesto definir expressamente um objetivo não-epistêmico.

TUE sobre a (ir)racionalidade de métodos de atualização não-Bayesianos. Num caminho parecido, Dantas (2016, cap. 5), investiga como um agente não-monotônico, que expressa a teoria do raciocínio anulável de Pollock (1995), lida com incerteza em comparação a um agente Bayesiano.

Na opção (iii), as hipóteses são derivadas de dados empíricos acerca de como humanos ou animais não-humanos raciocinam. Essa opção é viável porque, entre os agentes conhecidos por nós, humanos os que mais obviamente exibem racionalidade relativamente à maioria das tarefas. Em outras tarefas, porém pode acontecer de humanos serem superados por animais não-humanos (ex. HERBRANSON; SCHROEDER, 2010). Esses dados, porém, não indicam diretamente racionalidade porque o modo como humanos raciocinam exibe vieses e heurísticas, que *não necessariamente* (mas possivelmente) exibem racionalidade (cf. KAHNEMAN, 2012). De qualquer maneira, apesar de pouco explorada, esta opção é interessante por causa do princípio ‘*deve* (epistêmico) implica *pode* (cognitivo)’. Essa opção evita um problema comum em Epistemologia: normas de racionalidade impossíveis de serem satisfeitas por agentes com recursos cognitivos limitados (ex. o fechamento de crenças sob consequência lógica). Dantas (2016, cap. 5) parte da hipótese de Kahneman (2012) de que o raciocínio humano é realizados por dois sistemas (Teoria do Processo Dual)²¹ e da hipótese de Stenning e Lambalgen (2008) de processos de um desses sistemas usa lógicas não-monotônica para estudar a racionalidade destes processos (cf. Seção 2).

Simulação

Nessa fase do estudo, as informações sobre a classe de problemas e sobre as hipóteses a serem investigadas são implementadas numa simulação de computador. A viabilidade computacional dessa simulação depende da adoção, nas hipóteses e classes de problemas, de abstrações (a incorporação de alguns, mas não todos, parâmetros relevantes) e mesmo de idealizações (a distorção deliberada de parâmetros relevantes). Por outro lado, há o perigo de que a adoção de abstrações e idealizações falsifiquem o resultado. Nas

²¹ “O Sistema 1 opera automática e rapidamente, com pouco ou nenhum esforço e nenhuma percepção de controle voluntário. O Sistema 2 aloca atenção às atividades mentais laboriosas que o requisitam, incluindo cálculos complexos” (KAHNEMAN, 2012, p. 29).

seções anteriores, falamos algumas vezes de ‘modelos’ utilizados em EC (ex. HK, Laputa, etc). Um modelo é, em certo sentido, um pacote de abstrações e idealizações que se crê bem-sucedido. Por exemplo, em ambos HK e Laputa, mas não no modelo utilizado por Baumgaertner (2014), a localização espacial dos agentes é abstraída. Uma idealização adotada em quase todos estudos em EC é atualizações de crenças serem instantâneas²².

Na EC Individualista, a simulação precisa ser programa ‘diretamente’ utilizando alguma linguagem de programação (ex. LISP, R, Java, etc). Na EC social, existem plataformas com interface amigável para gerar simulações (ex. Laputa). Porém, a escolha de utilizar alguma dessas plataformas implica na escolha do modelo subjacente. As simulações usadas em EC são baseadas em agentes, no sentido de simularem o comportamento de indivíduos a partir de regras locais (relativas ao indivíduo), em vez de equações que regem todo o sistema (ex. como nas simulações utilizadas em meteorologia, cf. WINSBERG, 2019, sec. 2.2). Simulações baseadas em agentes são usadas em outras ciências (Ciências Sociais, ecologia, IA, etc). Isto relevante porque muitas vezes modelos utilizados em EC são adaptados destas disciplinas. Por exemplo, HK tem sua origem em Ciências Sociais, Netlogo é utilizado em ecologia e o modelo utilizado de Dantas (2016, cap. 5) tem origem em IA. É um bom caminho tentar adaptar para a EC modelos utilizados em outras ciências. Mas é preciso estar atento. Noutro sentido, um modelo é uma classe de classes de problemas (a classe das classes de problemas que aquele modelo é capaz de expressar). Então, a escolha do modelo implica numa limitação do estudo a algumas classes²³.

Um estudo consciencioso em EC precisa discutir as possíveis distorções advindas do modelo escolhido. Nesse sentido, uma das vantagens de adotar um modelo existente é que suas abstrações e idealizações são bem conhecidas e discutidas. Porém, mesmo nesse caso, é interessante procurar

²² Em Dantas (2016, cap. 5) é medida a quantidade de tempo necessária para cada atualização.

²³ Por exemplo, Hahn et al (2018) estudam como diferentes características estruturais das redes sociais afetam sua competência coletiva. Idealmente, eles deveriam fixar o valor para uma característica (ex. coeficiente de agrupamento), gerar aleatoriamente redes e tirar a média dos resultados. Eles poderiam variar o valor desta característica e ver como isto afeta a competência coletiva. Mas Laputa é incapaz de variar aleatoriamente a estrutura de grafos. Por isso, Hahn et al precisaram escolher (mais ou menos arbitrariamente) 36 redes para simular. Mesmo não invalidando os resultados, estes são uma generalização a partir de poucos casos (36).

validar o modelo escolhido a partir de dados empíricos²⁴. Olsson (2013), por exemplo, argumenta que as idealizações de Laputa são justificáveis do ponto de vista da Teoria do Processo Dual (cf. KAHNEMAN, 2012). Nesse modelo, agentes tomam os testemunhos de outros agentes sempre como probabilisticamente independentes (o que comumente não é o caso quando, por exemplo, a fonte da informação for a mesma). Porém, Olsson argumenta que essa idealização pode ser vista como uma estratégia plausível do Sistema 1, usada na ausência de motivos concretos de desconfiança.

Análise

Uma simulação em EC geralmente funciona como um ciclo (*loop*). O ciclo começa com o ambiente emitindo aos agentes informação sobre seu estado atual (ex. o que outros agentes estão lhe afirmando); os agentes, então, atualizam suas crenças e executam a quantidade raciocínio que for necessária para retornar uma ação para o ambiente (ex. afirmar algo). O ambiente, por fim, atualiza seu estado dadas as ações dos agentes e o ciclo é reiterado. O ciclo termina quando os objetivos do problema ou um limite de iterações preestabelecido forem atingidos. Os dados sobre a simulação para posterior análise são produzidos introduzindo testes (expressões condicionais do tipo ‘se-então’) e contadores (variáveis) dentro do ciclo. Por exemplo, podemos incluir um teste para checar, no começo de cada iteração, se o agente completou o objetivo não-epistêmico. Outro exemplo: podemos incluir testes e contadores para observar, em cada iteração do ciclo, o quanto o agente completou o objetivo epistêmico. No caso das plataformas de interface amigável, como o Laputa, os dados são coletados automaticamente. O lado ruim é que não é possível mudar o modo os cálculos são feitos ou colher informações diferentes das pré-programadas.

Num estudo em EC, é sempre necessário medir o quanto os agentes cumprem o objetivo epistêmico. Boa parte da literatura em EC é veritista. Nesse caso, o valor epistêmico é diretamente proporcional à quantidade de verdade e inversamente à de falsidade no conjunto de crenças dos agentes.

²⁴ Validar um modelo é o processo de determinar se o modelo escolhido é uma representação boa o suficiente do sistema real que está sendo simulado. No caso da EC, esse sistema são (grupos de) agentes racionais.

Existe uma literatura considerável sobre medidas desse tipo, mas uma medida de inacurácia muito comum é a medida de Brier (cf. TITELBAUM, no prelo, cap. 10). Porém, diferentemente do que a literatura em TUE leva a crer, esta não é a única medida possível. Por exemplo, simulações geradas em Laputa utilizam uma medida mais simples, o V-valor (cf. OLSSON, 2011). Um problema compartilhado pela medida de Brier e o V-valor é que essas medidas levam em conta apenas a quantidade de falsidade num conjunto de crenças, mas não a quantidade de verdade. Por isso, essas medidas só podem ser utilizadas para comparar agentes que tenham o mesmo número de crenças. Dantas (no prelo) propõe uma medida que permite comparar agentes com número diferente de crenças. Dantas (2016, cap. 5) utiliza essa medida (cf. Seção 2).

2. Mundo de Wumpus

Dantas (2016, cap. 5) compara como um agente Bayesiano e um agente não-monotônico lidam com raciocínio incerto²⁵ num ambiente mais realista que a média em EC. O ambiente é mais realista no sentido de que considera a consequência de ações que vão além da acurácia de crenças (ex. os agentes podem morrer). Esse estudo ainda não foi publicado.

São resultados importantes da TUE que conjuntos de crenças que expressam funções de probabilidade minimizam a inacurácia *esperada* em comparação com os que não o fazem (probabilismo) e que a atualização de crenças por condicionalização minimiza a inacurácia esperada das crenças resultantes em comparação com outros métodos de atualização (cf. TITELBAUM, no prelo, cap. 10). Porém, uma questão (também colocada por DOUVEN, 2013) é saber quão importante é minimizar a inacurácia *esperada*, por exemplo, para a realização de objetivos não-epistêmicos (ex. se manter vivo num ambiente hostil). Outro aspecto interessante desse estudo é que ele leva em consideração o fato de que agentes epistêmicos reais (ex. humanos) têm recursos cognitivos limitados (ex. memória limitada). Será que a condicionalização é a melhor estratégia de atualização de crenças para agentes com recursos cognitivos limitados num ambiente hostil?

²⁵ ‘Raciocínio incerto’, aqui, significa raciocínio cuja conclusão não se segue dedutivamente das premissas.

Problema

Para comparar os agentes Bayesiano e não-monotônico, é usada uma versão epistêmica do Mundo de Wumpus (MW), uma classe de problemas muito usado em IA para investigar questões sobre raciocínio incerto (cf. RUSSELL; NORVIG, 2010, p. 236). O ambiente do MW é uma caverna, que consiste em salas conectadas por galerias e cercadas por paredes. À espreita em algum lugar na caverna está o Wumpus, uma besta feroz que mata qualquer um que entre ali. O odor (fétido) do Wumpus pode ser sentido nas salas adjacentes àquela em que ele se encontra. Outras salas da caverna contêm poços profundos, que também matam qualquer um que entre ali. Nas salas adjacentes a um poço, é possível sentir uma leve brisa. O objetivo de um agente em MW é encontrar uma pilha de ouro, que está escondida em alguma sala da caverna, e sair vivo de lá.

Uma instância de MW pode ser representada como uma matriz, em que a posição (x, y) representa a sala na linha x e coluna y (cf. Figura 1). A entrada da caverna fica em $(0, 0)$. Ao explorar a caverna, o agente encontra brisas [odores] nas salas diretamente adjacentes às salas que contêm poços [Wumpus] e um brilho nas salas que contêm o ouro. O agente conhece seu estado inicial, objetivos, além das regras gerais do problema, mas começa em estado de ignorância sobre a localização dos poços, Wumpus e ouro.

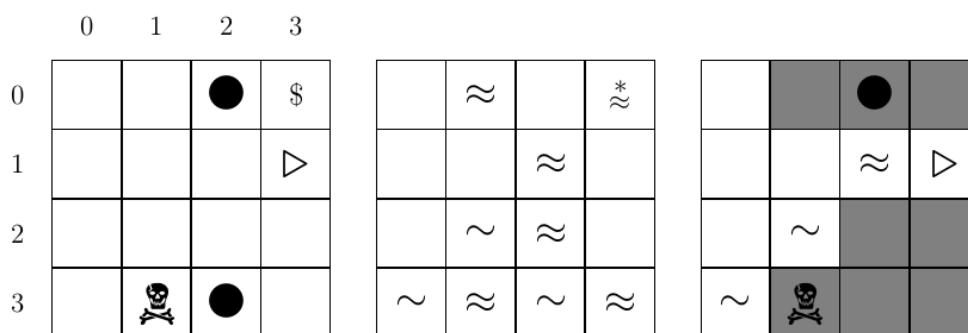


Figura 1: Da esquerda para a direita, mapa de uma instância de EMW, informações acessíveis, e características do mapa conhecidas pelo agente. Os símbolos significam o seguinte: ● = poço, ☠ = Wumpus, \$ = ouro, ▷ = agente (virado para a direita), ≈ = brisa, ~ = odor e * = brilho. Áreas visitadas estão em branco, áreas não-visitadas estão em cinza.

MW tem algumas características que justificam seu uso em investigações sobre raciocínio incerto. Por exemplo, agentes não têm informação

direta sobre a localização de poços [Wumpus], podendo apenas inferi-las a partir de informação indireta sobre brisas [odores]. Porém, os únicos objetivos do MW são não-epistêmicos: achar o outro e sair da caverna. Por isso, Dantas (2016, cap. 5) utiliza versão epistêmica de MW (MWE). Em MWE, além de achar o ouro e sair vivo, o agente tem como objetivo formar o conjunto de crenças mais compreensivo e acurado sobre a caverna (quais salas contêm poços, Wumpus, etc e quais não contêm). Além disso, há uma camada extra de incerteza: existe uma probabilidade (baixa) do agente ter falsos positivos persistentes sobre a localização de brisas [odores].

Hipóteses

O agente Bayesiano tem como características: (b1) suas crenças são modeladas como graus de crença com valores no intervalo $[0, 1]$; (b2) seus graus de crença expressam uma função de probabilidade; e (b3) seus graus de crenças são atualizados por condicionalização sobre a evidência disponível. Esse modelo é bem conhecido, por isso, não vou me alongar nele.

A lógica clássica é monotônica no seguinte sentido: uma conclusão adotada a partir de um conjunto de razões não é anulada por novas razões. Numa lógica não-monotônica, porém, isto ocorre. Stenning e van Lambalgen (2008) defendem que “partes relevantes dos processos do Sistema 1 são implementações bastante diretas de uma série de lógicas não-monotônicas” (ex. p. 130). Eles argumentam que a adoção um modelo baseado em lógica não-monotônica explicaria uma série de resultados empíricos da literatura sobre vieses e heurísticas. Por exemplo, tarefas de supressão (cf. BYRNE, 1989) sugerem que a adição de novas premissas pode fazer com que sujeitos suspendam conclusões derivadas por *modus ponens*, seriam facilmente explicados deste ponto de vista. Dantas (2016, cap. 5) pretende investigar a racionalidade desta estratégia de Sistema 1.

O agente não-monotônico tem as seguintes características gerais: (d1) suas crenças são modeladas como tendo dois valores possíveis (adotada e anulada); (d2) suas crenças são adotadas a partir da presença de razões e (d3) anuladas a partir da presença de anuladores.

Razões podem ser conclusivas ou anuláveis, dependendo do suporte epistêmicos que elas fornecem (cf. POLLOCK, 1995, p. 85). O suporte epis-

têmico de razões conclusivas não pode ser anulado por novas informações. Por exemplo, a ausência de brisa [odor] numa sala é razão conclusiva para a ausência de poço [Wumpus] nas salas adjacentes. Por outro lado, o suporte de razões anuláveis pode ser anulado por novas informações (anuladores). Por exemplo, a presença de brisa [odor] numa sala é razão anulável para a presença de poços [Wumpuses] nas salas adjacentes. O suporte epistêmico de uma razão anulável pode ser anulado por novas informações. Anuladores podem ser rebutadores ou solapadores. Rebutadores atacam a conclusão de uma razão (ex. por suportar conclusão contraditória). Por exemplo, a ausência de brisa [odor] em alguma sala adjacente rebuta qualquer razão para a conclusão de que naquela sala há um poço [Wumpus]. Um solapador ataca a conexão entre uma razão e sua conclusão. Por exemplo, uma razão para haver Wumpus noutra sala solapa qualquer razão para haver Wumpus numa sala (só há um Wumpus).

Dada a diferença entre (b1) e (d1), seria esperado que o agente Bayesiano exibisse uma acurácia maior em relação ao objetivo epistêmico. Afinal, graus de crença permitem ao agente Bayesiano fazer distinções mais sutis entre hipóteses. Além disso, os resultados da TUE sobre minimização da inacurácia esperada parecem implicar que o agente Bayesiano tenha graus de crenças mais acurados após sucessivas atualizações. Nesse contexto, parece haver pouca expectativa para a (maior) racionalidade da estratégia do Sistema 1 em MWE.

Simulação

O agente começa na sala (0, 0), olhando para a direita. Para cada sala, menos a (0, 0), há uma probabilidade de .2 de haver um poço. O Wumpus e o ouro são posicionados aleatoriamente (com distribuição uniforme) em uma sala diferente da (0, 0). Nas salas diretamente adjacentes àquelas que contém um poço [Wumpus] há uma brisa [odor]. Nas salas que contém ouro, há um brilho. Para cada sala, menos a (0, 0), há uma probabilidade independente de .01 de haver uma brisa [odor] (possibilidade de falso-positivo). O agente pode mover-se para a frente, girar 90° para a esquerda ou direita, pegar algo (se houver algo na sala) e sair da caverna quando está na sala (0, 0) com o ouro. O agente morre se ele entra numa sala contendo um

poço ou o Wumpus. A simulação termina quando o agente sai da caverna com o ouro ou quando ele morre.

O objetivo não-epistêmico do MWE é sair da caverna (vivo) com o ouro. Por isso, há uma pontuação de +1000 por estar vivo e +1000 por ter o ouro. Além disso, há uma penalidade de -1 para cada ação. O objetivo epistêmico do MWE é a adoção de um conjunto de crenças mais compreensivo e acurado possível sobre o interior da caverna. Por isso, há uma pontuação de $+ [1000 * (t - f) / (t + f + c)]$, em que t e f são (respectivamente) quantidade de verdade e de falsidade no conjunto de crenças do agente e c é uma constante (de fato, 0). A equação $(t - f) / (t + f + c)$ é discutida em Dantas (no prelo). Em caso de morte, todas as pontuações positivas são zeradas.

Na simulação, agentes são compostos por quatro módulos: percepção, memória, cognição prática e epistêmica. Os agentes Bayesianos e não-monotônicos compartilham a percepção, memória e parte considerável da cognição prática, mas têm cognição epistêmica baseadas nos modelos Bayesianos e não-monotônicos (respectivamente). O agente Bayesiano tem sua cognição epistêmica baseada em algoritmos conhecidos para cálculo de probabilidades a partir de redes Bayesianas (cf. PEARL, 1986). Um ponto importante é, pelo fato das redes Bayesianas para poços serem multiplamente conectadas, a atualização de graus de crença sobre a localização dos poços depende da construção de tabelas de probabilidade conjunta.

A cognição epistêmica do agente não-monotônico é mais peculiar. O agente apenas soma o número de razões disponíveis para a presença de poço [Wumpus] numa sala²⁶. Se o número de razões for positivo, ele adota a crença de que há um poço [Wumpus] naquela sala. Rebutadores (que, em MWE, são sempre conclusivos) diminuem o número de razões para um valor negativo muito grande. Um valor negativo para razões faz com que o agente adote a crença de que *não* há um poço [Wumpus, ouro] naquela sala. O caso do Wumpus é mais complexo porque usa solapadores. O agente pode ter um número positivo de razões para a presença de Wumpus em mais de uma sala. Mas o conhecimento de que existe apenas um Wumpus solapa estas ra-

²⁶ Ao fazer isso, o agente está assumindo que essas razões são independentes. Aqui, como em Olsson (2013), podemos dizer que esta idealização pode ser vista como uma estratégia plausível do Sistema 1, o que está em linha com o fato de estarmos testando uma estratégia de Sistema 1.

zões. Neste caso, o agente simplesmente compara o valor das razões e adota a crença de que há Wumpus na sala com valor maior. No caso de valores iguais, ele adota a crença disjuntiva de que o Wumpus está em alguma das salas com valor maior. Por mais simplório que esse algoritmo pareça, ele implementa a semântica da múltipla atribuição (POLLOCK, 1995, p. 308), com pequenas modificações.

Análise

A simulação foi rodada com cavernas de dimensões quadradas e com tamanhos de 2 a 10, com 1000 repetições para cada tamanho. Em cada repetição, é gerada uma caverna com as características descritas acima, com o tamanho especificado, mas com posições de poços, Wumpus e ouro aleatórias. Na Figura 2, são apresentadas a média entre as 1000 repetições.

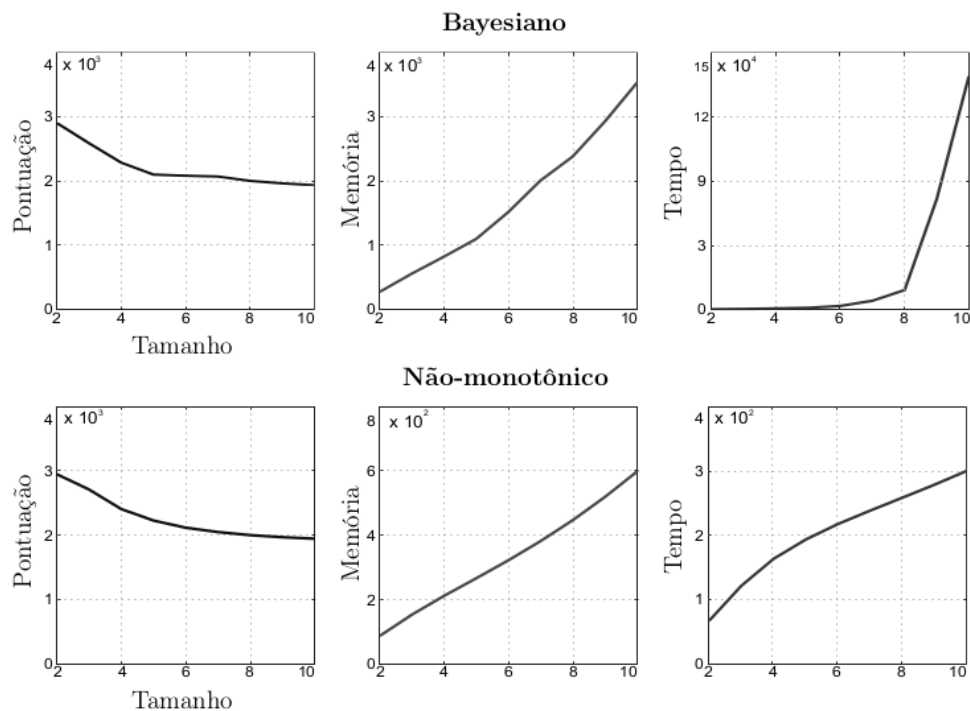


Figura 2: Resultados dos agentes Bayesiano e não-monotônico para cavernas de tamanho 2 a 10 (médias de 1000 repetições). Na esquerda, a pontuação de ambos estabiliza perto de 2000 (2/3 do máximo). No centro, ambos agentes são polinomiais em memória. Na direita, enquanto o agente Bayesiano é exponencial em tempo, o não-monotônico é polinomial.

Os agentes Bayesiano e não-monotônico têm pontuações semelhantes (Figura 2, esquerda). A pontuação de ambos estabiliza num valor próxi-

mo de 2000 (2/3 do máximo). O agente tende a não morrer em 2/3 das repetições, achar o outro em 2/3 das repetições e ter cinco vezes mais crenças verdadeiras que falsas $[(5f-f)/(5f+f) = 2/3]$. Esse é um bom resultado, uma vez que 21% das instâncias de MW são impossíveis de resolver (ex. o ouro está cercado de poços, cf. RUSSELL; NORVIG, 2010, p. 198).

Apesar de ser esperado que o agente Bayesiano fosse mais acurado, os resultados sugerem que o agente não-monotônico é tão acurado quanto o Bayesiano²⁷. Em casos como este, a racionalidade da adoção de uma ou outra estratégia epistêmica depende de outros fatores. Uma inovação de Dantas (2016, cap. 5) é a investigação da eficiência computacional de destas estratégias. A motivação é seguinte: a Epistemologia (humana) está diretamente interessada numa noção de racionalidade para sujeitos com recursos cognitivos limitados (afinal, humanos têm recursos cognitivos limitados). Quando lidamos com sujeitos com recursos cognitivos limitados, a noção de racionalidade parece estar (também) relacionada ao bom uso de recursos cognitivos escassos.

Dantas (2017) argumenta que a eficiência computacional de uma prática epistêmica (cf. SIPSER, 2012) é um parâmetro relevante para uma noção de racionalidade. A ideia do argumento é que um sujeito com recursos cognitivos limitados só poderia aproximar um ideal de racionalidade no limite de sucessivas iterações de uma prática epistêmica. Porém, estes sujeitos sempre param em algum ponto da iteração por falta de recursos cognitivos. A eficiência de uma prática epistêmica (individualista) estaria relacionada à quantidade de informação que precisa ser armazenada em cada iteração (memória) e ao número relativo de passos inferenciais utilizados em cada iteração (tempo). Nesse caso, um sujeito que execute uma prática epistêmica com requisitos de memória ou tempo que crescem exponencialmente tende a parar ‘muito antes’ que outro que utilize uma prática polinomial. Como algoritmos exponenciais estão relacionadas à força-bruta e algoritmos polinomiais com entendimento profundo de um problema (SIPSER, 2012, p. 285),

²⁷ Essa conclusão, porém, fica prejudicada pelo fato de eu ter colocado a pontuação em relação aos objetivos epistêmico e não-epistêmico no mesmo gráfico. Mesmo assim, acho que existam apenas duas opções: ou o agente Bayesiano performa melhor no objetivo epistêmico (e pior no não-epistêmico) ou ambos performam igualmente nos dois objetivos (não considero a terceira alternativa por causas dos resultados da TUE). Em ambos os casos, creio que a argumentação de minhas conclusões se segue.

este argumento seria uma defesa do seguinte princípio epistêmico: ‘não há racionalidade por força-bruta’.

O uso de memória foi medido em número de bytes armazenados. O uso de memória de ambos os agentes cresce polinomialmente em relação ao tamanho da caverna (Figura 2, centro). Por outro lado, há uma diferença relevante entre esses agentes no uso de tempo. Enquanto o tempo utilizado pelo agente Bayesiano cresce exponencialmente em relação ao tamanho da caverna, o tempo do agente não-monotônico cresce apenas polinomialmente (Figura 2, direita). O tempo utilizado foi medido em termos de quantidade de passos inferenciais utilizados nas atualizações de crenças (ex. a mudança do valor a uma variável). O agente Bayesiano tem crescimento exponencial em tempo por causa da necessidade de usar tabelas de probabilidade conjunta para atualizar os graus de crença sobre poços²⁸.

Esses resultados, porém, devem ser interpretados cuidadosamente. Os resultados dependem de duas características de MWE: (i) a prevalência maior de raciocínio diagnóstico (e não causal)²⁹ e (ii) a possibilidade de ignorar quase completamente relações mais complexas de dependência condicional. Raciocínio diagnóstico é mais difícil para agentes Bayesiano porque probabilidades condicionais são mais naturalmente expressas como a probabilidade de um efeito dadas suas causas e não de causas dados seus possíveis efeitos (cf. PEARL, 1988, p. 35). O modelo não-monotônico, por outro lado, é indiferente a esta distinção (razões podem ir de efeitos para causas e vice-versa). De fato, dado o enunciado de MWE, é mais natural que razões funcionem na direção efeitos-causas.

A segunda característica do MWE que favorece o agente não-monotônico é a possibilidade de ignorar relações mais complexas de dependência condicional. O agente Bayesiano exhibe ‘naturalmente’ um tipo de raciocínio chamado de ‘explaining away’ (PEARL, 1988, p. 7). Por exemplo, se há duas causas possíveis para um efeito observado, a descoberta de que uma das causas ocorre diminui a probabilidade de que outra causa ocorra. O

²⁸ É bem sabido que tabelas de verdade (e, conseqüentemente, tabelas de probabilidade conjunta) crescem exponencialmente dado o número de variáveis consideradas (ex. salas que podem conter poços).

²⁹ Raciocínio diagnóstico vai de um efeito para sua causa (ex. de brisas para poços); raciocínio causal vai de uma causa para seu efeito (ex. de poços para brisas).

agente não-monotônico apenas pode apenas imitar este tipo de raciocínio usando solapadores. O número de solapadores poderia crescer exponencialmente em relação ao tamanho da caverna porque, no limite, a ocorrência de um poço [Wumpus] em cada uma das salas poderia ser um solapador para razões para crer que existem poços [Wumpus] em cada uma das outras salas. Porém, isso não acontece porque, ao raciocinar sobre poços, o agente não-monotônico simplesmente ignora qualquer relação de dependência condicional (qualquer razão mínima para crer que há um poço uma sala é razão suficiente para crer que há um poço naquela sala). Ao raciocinar sobre Wumpus, isso não pode ser feito porque há somente um Wumpus. Porém, nesse caso, o agente não-monotônico assegura seu caráter polinomial em tempo porque raciocina simplesmente somando e diminuindo razões.

Em suma, diferentemente do que era esperado inicialmente, os resultados das simulações mostram que os agentes Bayesiano e não-monotônico são igualmente acurados em casos de MWE (apesar da nota-de-rodapé 26). Além disso, os resultados sugerem que práticas epistêmicas baseadas em lógicas não-monotônicas tendem a ser computacionalmente mais eficientes que aquelas baseadas em probabilidades que em situações de raciocínio sob incerteza com as características do MWE (maior prevalência de raciocínio diagnóstico e possibilidade de ignorar relações de dependência condicional). Isso está em linha com a hipótese de Stenning e van Lambalgen de que processos de Sistema 1 implementam lógicas não-monotônicas, especialmente se levarmos em conta as noções de Sistema 1 e 2:

O Sistema 1 opera automática e rapidamente, com pouco ou nenhum esforço e nenhuma percepção de controle voluntário. O Sistema 2 aloca atenção às atividades mentais laboriosas que o requisitam, incluindo cálculos complexos (KAHNEMAN, 2012, p. 29).

De fato, o suposto processo de Sistema 1 (o raciocínio não-monotônico) exige menos esforço que o processo de Sistema 2 (o cálculo de probabilidades). Os resultados das simulações sugerem, porém, sugerem uma conclusão mais importante. Eles demonstrariam que, ao menos numa classe de problemas, o processo de Sistema 1 é mais racional. Este tipo de resultado parece apontar para uma noção de ‘racionalidade’ que não colapse para um diagnóstico de irracionalidade geral frente os resultados da literatura so-

bre heurísticas e vieses. Nas palavras de Kahneman: “Nossos pensamentos e ações são rotineiramente guiados pelo Sistema 1 e em geral são corretos [ao menos, racionais]” (2012, p. 520).

Considerações Finais

Não restou muito a ser dito nesta conclusão, além de uma provocação. Ashton e Mizrahi (2018) analisaram 478.719 artigos de Filosofia publicados entre 1840 e 2012 e indexados no JSTOR Data for Research corpus (df.jstor.org) à procura de palavras indicativas de argumentação dedutiva e indutiva. Apesar das limitações metodológicas (palavras indicativas não são indicadores conclusivos do tipo de argumentos usado), os resultados indicam uma interessante tendência em Filosofia. Os resultados sugerem que a maioria dos artigos filosóficos ainda usam argumentos dedutivos. Porém, eles também sugerem que a diferença entre a porcentagem dos artigos propondo argumentos dedutivos e indutivos está diminuindo com o passar do tempo. Ou seja, a argumentação dedutiva estaria perdendo seu *status* como a forma dominante de argumentação filosófica. Faça parte dessa tendência!

Referências

- ANGERE, Staffan. The defeasible nature of coherentist justification. *Synthese*, v. 157, p. 321-335, 2007.
- BAUMGAERTNER, Bert. Yes, no, maybe so: a veritistic approach to echo chambers using a trichotomous belief model. *Synthese*, v. 191, n. 11, p. 2549-2569, 2014.
- BLOWN, Peter; BUCKWALTER, Wesley; TURRI, John. Gettier Cases: A Taxonomy. In: BORGES, R. et al. (eds.). *Explaining Knowledge: new essays on the Gettier problem*. Nova Iorque: Oxford University Press, 2018.
- BYRNE, Ruth. Suppressing valid inferences with conditionals. *Cognition*, v. 31, p. 61-83, 1989.
- COMESAÑA, Juan. Unsafe knowledge. *Synthese*, v. 146, n. 3, p. 395-404, 2005
- CROSSCOMBE, Michael; LAWRY, Jonathan. A model of multi-agent consensus for vague and uncertain beliefs. *Adaptive Behavior*, v. 24, n. 4, p. 249-260, 2016.

CUMMINS, Robert. Reflection on reflective equilibrium. In: DEPAUL, M.; RAMSEY, W. (eds.). *Rethinking intuition*, Lanham: Rowman & Littlefield, p. 113-128, 1998.

DANTAS, Danilo. How to (blind)spot the truth. No prelo. Disponível em: <academia.edu/35240910/How_to_Blind_Spot_the_Truth>.

DANTAS, Danilo. No rationality through brute-force. *Philosophy South (Filosofia Unisinos)*, v. 18, n. 3, p. 197-200, 2017.

DANTAS, Danilo. *Almost ideal: Computational Epistemology and the limits of rationality for finite reasoners*. Tese (Doutorado em Filosofia) - University of California, Davis, 2016.

DOUVEN, Igor. Computational models in Social Epistemology. In: FRICKER, M. et al (eds.). *The Routledge handbook of Social Epistemology*. Nova Iorque: Routledge, p. 457-465, 2020.

DOUVEN, Igor. Inference to the best explanation, Dutch books, and inaccuracy minimisation. *The Philosophical Quarterly*, v. 63, n. 252, p. 428-444, 2013.

DOUVEN, Igor. Simulating peer disagreements. *Studies in History and Philosophy of Science*, 41, n.2, p. 148-157, 2010.

FOLEY, Richard. Quine and Naturalized Epistemology. *Midwest Studies in Philosophy*, v. 19, p. 243-260, 1994.

FRANKFURT, Harry. *Sobre falar merda*. Rio de Janeiro: Editora Intrínseca Ltda, 2005.

GLASS, David. Coherence, explanation, and hypothesis selection. *The British Journal for the Philosophy of Science*, axy063, 2018.

GLASS, David. Inference to the best explanation: does it track truth? *Synthese*, v. 185, n. 3, p. 411-427, 2012.

GLASS, David. Coherence, explanation and Bayesian networks. In: O'Neill, M. et al. (eds.). *Artificial intelligence and cognitive science*. Berlin: Springer, p. 177-182, 2002.

GOLDMAN, Alvin. *Knowledge in a social world*. Clarendon Press, 1999.

GRIM, Patrick; SINGER, Daniel. Computational Philosophy. In: ZALTA, E. (ed.). *The Stanford encyclopedia of Philosophy*, 2020. Disponível em: <plato.stanford.edu/archives/spr2020/entries/computational-philosophy/>.

HAHN, Ulrike; HANSEN, Jens; OLSSON, Erik. Truth tracking performance of social networks: how connectivity and clustering can make groups less competent. *Synthese*, v. 197, n. 4, p. 1511-1541, 2018.

HEGSELMANN, Rainer; KRAUSE, Ulrich. Opinion dynamics and bounded confidence: models, analysis, and simulations. *Journal of Artificial Societies and Social Simulation*, v. 5, n. 3, 2002. Disponível em: <http://jasss.soc.surrey.ac.uk/5/3/2.html>.

_____. Truth and cognitive division of labor: first steps towards a computer aided Social Epistemology. *Journal of Artificial Societies and Social Simulation*, v. 9, n. 3, 2006. Disponível em: <http://jasss.soc.surrey.ac.uk/9/3/10.html>.

HERBRANSON, Walter; SCHROEDER, Julia. Are birds smarter than mathematicians? Pigeons (*Columba livia*) perform optimally on a version of the Monty Hall Dilemma. *Journal of Comparative Psychology*, v. 124, n. 1, p. 1-13, 2010.

KAHNEMAN, Daniel. *Rápido e devagar: duas formas de pensar*. São Paulo: Objetiva, 2012.

KIM, Jaegwon. What is ‘Naturalized Epistemology’? *Philosophical Perspectives*, v. 2, p. 381-405, 1988.

KIM, Minsun; YUAN, Yuan. No cross-cultural differences in the Gettier car case intuition: a replication study of Weinberg et al. 2001, *Episteme*, v. 12, n. 3, p. 355-361, 2015.

KNOBE, Joshua; NICHOLS, Shaun. Experimental Philosophy. In: ZALTA, E. (ed.). *The Stanford encyclopedia of Philosophy*, 2017. Disponível em: <http://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/>.

KORNBLITH, Hilary. *Knowledge and its place in nature*. Nova Iorque: Oxford University Press, 2002.

OLSSON, Eric. A simulation approach to veritistic social epistemology. *Episteme*, v. 8, n 2, p. 127-143, 2011.

OLSSON, Eric. (2013). A bayesian simulation model of group deliberation and polarization. In: ZENKER, F. (ed.). *Bayesian argumentation*. Nova Iorque: Springer, p. 113–134, 2013.

OLSSON, Eric. What is the problem of coherence and truth? *Journal of Philosophy*, v. 99, p. 246-272, 2002.

PALLAVICINI, Josefine; HALLSSON, Bjørn; KAPPEL, Klemens. Polarization in groups of Bayesian agents. *Synthese*, 2018.

PEARL, Judea. *Probabilistic reasoning in intelligent systems: networks of plausible inference*, São Francisco: Morgan Kaufmann Publishers Inc, 1988.

PEARL, Judea. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, v. 29, n. 3, p. 241-288, 1986.

POLLOCK, John. Procedural Epistemology: at the interface of Philosophy and AI. In: GRECO, J.; SOSA, E. (eds.). **The Blackwell Guide to Epistemology**. Nova Iorque: Blackwell Publishing, p. 383-414, 1999.

POLLOCK, John. *Cognitive carpentry: A blueprint for how to build a person*. Cambridge: The MIT Press, 1995.

POLLOCK, John. *How to build a person: a prolegomenon*. Cambridge: The MIT Press, 1989.

QUINE, W. V. O. Reply to Morton White. In: SCHILPP, P.; HAHN, L. (eds.). *The Philosophy of W. V. Quine*. Chicago: Library of Living Philosophers, v. 18, p. 663-665, 1986.

QUINE, W. V. O. A Epistemologia Naturalizada (trad. Carrilho, M.), 1969. In: CARRILHO, M.; SÀÁGUA, J. (eds.). *Epistemologia: posições e críticas*. Lisboa: Fundação Gulbenkian, p. 267-298, 1991.

RUSSELL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. Upper Saddle River: Prentice-Hall, 2010.

RYSIEW, Patrick. (2020), Naturalism in Epistemology. In: ZALTA, E. (ed.). *The Stanford encyclopedia of Philosophy*. URL = <<https://plato.stanford.edu/archives/spr2020/entries/epistemology-naturalized/>>, 2020.

SEYEDSAYAMDOST, Hamid. On normativity and epistemic intuitions: failure of replication. *Episteme*, v. 12, n. 1, p. 95-116, 2015.

SHAH, Nishi. How truth governs belief. *The Philosophical Review*, v. 112, n. 4, p. 447-482, 2003.

SIPSER, Michael. *Introduction to the Theory of Computation*. Stamford: Cengage Learning, 2012.

SOSA, Ernest. A defense of the use of intuitions in Philosophy. In: MURPHY, D.; BISHOP, M. (eds.). *Stich and his critics*, Malden: Wiley Blackwell, p. 101-112, 2009.

STENNING, Keith; VAN LAMBALGEN, Michiel. *Human reasoning and Cognitive Science*. Cambridge: The MIT Press, 2008.

STICH, Stephen. Naturalizing Epistemology: Quine, Simon and the prospects for Pragmatism. In: HOOKWAY, C.; PETERSON, D. (eds.). *Philosophy and Cognitive Science*. Cambridge: Cambridge University Press, p. 1-17, 1993.

TITELBAUM, Michael. *Fundamentals of Bayesian Epistemology*. No prelo. Disponível em: <tinyurl.com/ybfesffu>.

TRPIN, Borut; PELLERT, Max. Inference to the best explanation in uncertain evidential situations. *The British Journal for the Philosophy of Science*, v. 70, n. 4, p. 977-1001, 2018.

WEINBERG, Jonathan; NICHOLS, Shaun; STICH, Stephen. Normativity and epistemic intuitions. *Philosophical Topics*, v. 29, n. 1, p. 429-460, 2001.

WILLIAMSON, Timothy. *Knowledge and its Limits*. Nova Iorque: Oxford University Press, 2000.

WINSBERG, Eric. Computer simulations in Science. In: ZALTA, E. (ed.). *The Stanford encyclopedia of Philosophy*. 2019. Disponível em: <plato.stanford.edu/archives/win2019/entries/simulations-science/>.

WRENN, Chase. Epistemology as Engineering? *Theoria*, v. 72, n. 1, p. 60-79, 2006.

ZOLLMAN, Kevin. The communication structure of epistemic communities. *Philosophy of Science*, v. 74, n. 5, p. 574-587, 2007.