

PROJETO DE PESQUISA

SANIDADE, RAZÃO E FINITUDE

Danilo Fraga Dantas

danilo.dantas@academico.ufpb.br

Introdução

Um ser onisciente não precisa (talvez, nem deva) suspender o julgamento sobre qualquer questão porque ele sabe o valor de verdade de toda proposição. Um ser onipotente, com recursos cognitivos ilimitados, não precisa (talvez, nem deva) esquecer qualquer registro armazenado. Afinal, ela tem um espaço ilimitado na memória e um poder computacional ilimitado para lidar com qualquer quantidade de informação recuperada. Pela mesma razão, ela não tem problemas em entulhar sua mente com informação irrelevante. Nós não somos esse ser! Um fato fundamental sobre nossa situação cognitiva é que “os seres humanos estão na situação finita (*finitary predicament*) de ter limites fixos em suas capacidades cognitivas e no tempo que eles têm disponível” (Cherniak, 1986, p. 8). A racionalidade epistêmica (no que se segue, ‘racionalidade’) parece exigir que os ‘raciocinadores finitos’ (aqueles que, como nós, estão na situação finita) exibam um tipo de ‘parcimônia cognitiva’: a utilização eficiente de seus poucos recursos cognitivos.

Este fato é reconhecido nas Ciências Cognitivas, onde a “tese da cognição tratável” (van Rooij, 2008) é muitas vezes utilizada para restringir o espaço das teorias da racionalidade em nível computacional¹ (ex. Oaksford & Chater, 2007, p. 35). Os epistemólogos, porém, costumam descartar a eficiência cognitiva como sendo um valor meramente pragmático e defendem que a Epistemologia deva se preocupar somente com ‘valor epistêmico’. Diferentes propriedades (de conjuntos de crenças) podem ser consideradas fontes de valor epistêmico: fechamento lógico, consistência lógica, coerência (dadas as diferentes noções de coerência), suporte evidencial, etc. Nas últimas décadas, porém, alguns epistemólogos têm argumentado em favor do “veritismo”, a tese de que o valor epistêmico fundamental é a posse de crenças verdadeiras, mas não de crenças falsas (ex. Pettigrew, 2019, p. 761)².

¹ Refiro-me ao segundo dos três níveis de explicação de fenômenos cognitivos propostos por Marr (1982).

² “A fonte fundamental de valor epistêmico para um estado doxástico é a medida em que ele representa o mundo corretamente: isto é, seu valor epistêmico fundamental é determinado inteiramente por sua verdade ou falsidade. ...Qualquer outra fonte de valor para uma crença é derivada, não fundamental: isto é, ela é valiosa na medida em que promove ou de outra forma serve à fonte fundamental” (Pettigrew, 2019, p. 761).

QUESTÃO: supondo que o valor epistêmico é veritístico, como justificar a tese de que a racionalidade exige um tipo de parcimônia cognitiva dos raciocinadores finitos?

Em Dantas (2022), proponho a noção de sanidade epistêmica, uma propriedade de parcimônia entre a manutenção de crenças verdadeiras e não falsas, e a consideração de nossas limitações cognitivas. ‘Sanidade’ é muitas vezes utilizado para atribuir a ausência de doenças psiquiátricas. Nesse caso, a ‘doença epistêmica’ paradigmática seria a tendência a manter crenças flagrantemente falsas e injustificadas (ex. teóricos da conspiração). Mas ‘sanidade’ também tem o significado de ‘saúde mental’ e este é o significado que estou explorando para designar se um raciocinador finito está ‘em boa forma’ (*fitness*) para alcançar valor epistêmico (crenças verdadeiras). Seja o ‘potencial alético’ de um agente a quantidade de valor epistêmico *extra* que é esperado que ele alcance a partir da exploração do ambiente, suas crenças e capacidades de raciocínio atuais. A sanidade epistêmica de um agente estaria relacionada à manutenção de crenças que aumentam seu potencial alético (crenças ‘relevantes’), mas não daquelas que o diminuem. A racionalidade, então, exigiria de raciocinadores finitos que eles mantenham sua sanidade epistêmica, evitando assim a manutenção de crenças não-relevantes. O resultado é um veritismo impuro, em que é relevante como o agente reage à evidência.

O objetivo desse projeto é utilizar as noções de sanidade epistêmica, potencial alético e relevância para investigar assuntos relacionados a parcimônia cognitiva, entre eles a evitação de entulho (*clutter avoidance*), o esquecimento, a suspensão de julgamento e a divisão social do trabalho cognitivo. Penso que esses são os principais contribuidores para nossa sanidade epistêmica. Esse projeto é a continuação de investigações que venho desenvolvendo desde meu doutorado e que já renderam alguns frutos, como alguns artigos que serão relevantes na implementação do projeto. Entre eles, Dantas (2017), que ganhou o prêmio *Philosophy South prize*, e Dantas (2021) e Dantas (2022), que foram publicados em revistas internacionais de ponta (*Erkenntnis* e *Episteme*). Por fim, a parte desse projeto sobre esquecimento foi agraciada com o *Seal of Excellence* da European Commission.

1. Evitação de entulho

Harman (1986, p. 12) expressa assim seu princípio da evitação de entulho (CA): “não devemos entulhar nossa mente com trivialidades”, em que seus principais exemplos de

“trivialidades” são algumas verdades lógicas e consequências lógicas de nossas crenças³. Em outros lugares, Harman (1986) nos alerta para não entulhar nossas mentes com assuntos “sem utilidade” (p. 7), “sem sentido” (p. 12), “sem importância” (p. 42) e “nos quais não temos interesse” (p. 55). Mas o que torna um assunto interessante para alguém? A resposta mais comum faz do interesse um valor subjetivo e pragmático: “grosso modo, é esse alguém ter algum desejo ou interesse atendido por ter crenças no assunto em questão” (Friedman, 2018, p. 3). Essa noção de interesse faz do CA um princípio de uso muito restrito para a Epistemologia (cf. Friedman, 2018, p. 9). Por exemplo, um teórico da conspiração certamente acha sua teoria da conspiração favorita interessante nesse sentido, porém esse é exatamente o tipo de assunto com o qual devemos evitar entulhar nossa mente. O CA seria mais interessante para a Epistemologia se fosse interpretado usando uma noção objetiva de interesse.

Há uma dica sobre como perseguir essa noção na literatura sobre memória. A teoria adaptativa da memória afirma que o desempenho de um sistema de memória deve ser entendido em termos de sua adaptação à estrutura do ambiente (Anderson, 1991). Um sistema de memória adaptativo recuperaria um registro somente se o custo cognitivo da recuperação não exceder o ganho esperado de valor prático (mas podemos adaptar a ideia para o valor epistêmico), o que depende da estrutura do ambiente. Foi essa dica que persegui na proposta de entender a noção de sanidade epistêmica em termos de potencial alético (Dantas, 2022). O interesse que uma crença para um agente estaria relacionado a quanto sua adoção contribuiria para seu potencial alético. Esta seria uma noção objetiva de interesse (no sentido de Goldman, 1999a, p. 9), que chamei de ‘relevância’ (epistêmica). Utilizando essa noção, o CA poderia ser entendido assim: um agente deve acreditar que ϕ somente se ϕ for relevante (para ele). Ou seja, o CA deixaria de ser um princípio escanteado em Epistemologia e passaria a ser central para o estudo da sanidade epistêmica.

São estas algumas questões relacionadas ao CA que pretendo investigar:

- a) Já que o potencial alético depende do valor epistêmico que um agente pode alcançar num determinado período e estamos lidando com raciocinadores finitos (com capacidades computacionais limitadas), precisamos considerar a eficiência

³ Exemplos de consequências lógicas “triviais” seriam disjunções “longas, estranhas, incongruentes (em termos de assunto), em que pelo menos um disjuncto já é conhecido ou pelo menos acreditados pelo agente, e condicionais cujos consequentes já são acreditados e cujos antecedentes são conjunções e disjunções longas, estranhas e incongruentes, e assim por diante” (Friedman, 2018, p. 3).

epistêmica dos padrões de raciocínio dos agentes. Essa questão é discutida em Dantas (2017, manuscrito) e Griffiths et al (2015), mas pretendo retomá-la.

- b) Já que o potencial alético depende de propriedades do ambiente (ex. a disponibilidade da evidência), vale a pena discutir se a noção de racionalidade que emerge da noção de sanidade epistêmica não é uma de racionalidade ecológica (Gigerenzer, 2021) e quais as consequências desta noção para a Epistemologia. Também são relevantes aqui questões sobre cognição adaptativa (Anderson, 1991).
- c) Seria interessante analisar a relação entre as noções epistêmica e pragmática (mais comum) de relevância. Por exemplo, as limitações cognitivas e o frame problem (Shanahan, 2016) podem levar o agente a priorização de evidências por questões pragmáticas. Seria esse comportamento racional? Se sim, isso levaria a uma infiltração pragmática (*pragmatic encroachment*) na noção de racionalidade (epistêmica) resultante (cf. Gao, 2021, sobre esse tipo de infiltração pragmática).

2. Esquecimento

Os epistemólogos prestam pouca atenção ao esquecimento. A posição padrão é a de que o esquecimento é algo ruim do ponto de vista epistêmico, exceto talvez quando alguém esquece informação que não mais aceita. Essa posição decorre da ideia de que o papel da memória é simplesmente o de armazenar informação adquirida no passado e de disponibilizá-la para uso futuro. Porém, atualmente há um consenso crescente de que a memória tem um papel ativo no processamento de informação, no pensamento hipotético (De Brigard, 2013), planejamento orientado ao futuro (Klein et al, 2009), etc. Nesse contexto, Andonovski (2020) propõe que a memória deve ser entendida como uma “faculdade de triagem”, cujo papel é o de resolver um problema de relevância: tornar disponível a informação “correta”, dadas importantes restrições de tempo e recursos cognitivos. Essa concepção aproxima o estudo da memória do da sanidade epistêmica.

Pretendo seguir a trilha de Michaelian (2011), que argumenta que o esquecimento virtuoso aumenta a confiabilidade de um sistema de memória. Considere um sujeito cujos demais sistemas cognitivos são confiáveis, de modo que os registros armazenados em sua memória sejam, em sua maioria, acurados no momento em que foram armazenados. À medida que o ambiente muda ao redor do sujeito, alguns dos registros antes acurados se tornarão inacurados. Quanto mais antigo um registro, maior é a chance dele ter se tornado inacurado. Nesse contexto, Michaelian afirma que é virtuoso o

esquecimento de registros “antigos”, no sentido de não terem sido acessados recentemente. Registros “antigos” geralmente não são relevantes, no sentido de serem improváveis de serem usados em inferências a partir da evidência presente no ambiente. Michaelian propõe o princípio da eliminação de entulho (CE): “se a mente de alguém está cheia de trivialidades, ele deve removê-las”. O CE seria a contrapositiva do CA: se a crença de que ϕ não é relevante, então você não deve acreditar que ϕ (ex. você deve esquecer de que ϕ).

São estas algumas questões relacionadas ao esquecimento que pretendo investigar:

- a) A discussão filosófica sobre a relação entre crença e memória (semântica quanto episódica), em geral, não parece levar em conta a literatura empírica relevante (cf. Porot & Mandelbaum, 2021). Por isso, seria interessante analisar como a posse de crenças (explícitas e implícitas) depende da memória. Penso que a relevância de crenças explícitas poderia ser entendida a partir do modo como elas ‘encapsulam’ crenças implícitas (especialmente, crenças acessíveis, cf. Dantas, 2019, p. 39).
- b) Diferentes padrões de esquecimento afetam diferentemente a sanidade epistêmica. São diferentes padrões de esquecimento: esquecer aleatoriamente, apenas os registros “antigos” ou apenas os que não coerem com nossas crenças (cf. Singer et al, 2021). Pretendo investigar a eficácia de diferentes padrões de esquecimento em diferentes tipos de ambiente (ex. ambientes que mudam abruptamente ou gradualmente) e como eles interagem com diferentes padrões de raciocínio.

3. Suspensão de julgamento

Os epistemólogos geralmente distinguem três atitudes doxásticas categoriais: um agente pode acreditar ou desacreditar uma proposição, mas pode também manter uma atitude de neutralidade comprometida em relação a ela (‘suspensão de julgamento’). A mera ausência de crença e descrença não é suficiente para a suspensão, pois alguém que nunca considerou uma proposição não mantém uma atitude doxástica em relação a ela (Friedman, 2013, p. 167). Suspensão é uma atitude de neutralidade “comprometida” (Sturgeon, 2010, p. 133), no sentido que deve poder ser adotada e abandonada a partir de razões. Em Dantas (2022), argumento que a suspensão racional é uma virtude epistêmica entre os vícios da opinioação (falta de suspensão) e do ceticismo (excesso de suspensão) porque suspender na ausência de evidências é a atitude epistemicamente sã: é cognitivamente parcimonioso e evita a adoção de crenças não-relevantes (como as resultantes do princípio de indiferença).

Existem diferentes pontos de vista sobre a suspensão na literatura. A visão credal (ex. Sturgeon, 2008) afirma que suspender sobre uma proposição é manter crenças 'medianas' (*middling credences*) sobre ela e sua negação (isto é, crenças que não ultrapassam os limites lockeanos para crença e descrença). A visão de segunda-ordem (ex. Raleigh, 2019) afirma que suspender sobre uma proposição envolve não manter crenças de primeira-ordem sobre ela ou sua negação e manter uma crença de segunda-ordem sobre sua situação epistêmica em relação àquele assunto (ex. que você não está na posição de ter conhecimento sobre aquele assunto). A visão interrogativa (Friedman, 2015) afirma que suspender sobre uma proposição envolve adotar uma atitude interrogativa sobre ela: investigar ativamente sobre sua verdade ou falsidade. A visão anti-interrogativa (Lord, 2020) afirma que suspender sobre uma proposição envolve adotar uma atitude anti-interrogativa sobre ela: ignorar as evidências sobre sua verdade ou falsidade. Penso que devemos admitir dois tipos de suspensão, identificadas pelas visões interrogativa e anti-interrogativa, por mais que seja estranho que duas atitudes aparentemente opostas sejam “do mesmo tipo” (creio que esse ponto pode ser explicado, veja adiante).

São estas algumas questões relacionadas à suspensão que pretendo investigar:

- a) Creio que é possível explicar a relação entre as quatro visões da suspensão a partir de normas zetéticas, sobre quando devemos procurar ou ignorar evidência (cf. Friedman 2020), em termos de relevância. Resumidamente, devemos suspender quando as evidências sobre ambas uma proposição e sua negação são relevantes ou quando nenhuma o é. Essas condições estariam relacionadas aos limites lockeanos.
- b) Explicar exatamente como a suspensão pode colaborar para nossa sanidade epistêmica. A investigação implica discutir qual medida de valor epistêmico é adequada para avaliar os raciocinadores finitos e analisar a literatura empírica para entender em que situações suspender pode ser cognitivamente parcimonioso. Essa investigação já obteve algum avanço (c.f. Dantas, 2022, manuscrito e Métodos).

4. Epistemologia Social

Existem estudos de Epistemologia Social que partem do veritismo (cf. Goldman, 1999b, cap. 3). Alguns desses estudos discutem o papel de alguns dos principais contribuintes para nossa sanidade epistêmica para a competência coletiva de grupos de agentes. Por exemplo, Singer et al (2019) analisam as consequências de diferentes estratégias de esquecimento para a competência coletiva de grupos de agentes. Seguindo essa direção, identifiquei

algumas questões acerca da sanidade epistêmica que pode ter interesse tanto para a Epistemologia individualista quanto para a Social:

- a) A sanidade epistêmica parece pressupor que agentes abdicuem de ter crenças em diversas áreas do conhecimento quando podem confiavelmente adquirir crenças através do testemunho de especialistas. Isso serviria como uma justificação da divisão de trabalho cognitivo a partir de noções da Epistemologia Individualista. Seria também interessante saber quais estratégias de divisão do trabalho são mais eficientes para grupos. Por exemplo, é mais eficiente que cada indivíduo se especialize numa área ou algum grau de polimatia é salutar? É mais eficiente redes em que todos acessem a todos ou a existência de *gatekeeping* seria mais salutar?
- b) Também é uma questão saber se o conceito de sanidade epistêmica se aplica a grupos de indivíduos. Grupos de raciocinadores finitos têm limitações cognitivas, mas essas parecem ser menos severas que as de um indivíduo isolado. Nesse sentido, será que os principais contribuidores para a sanidade epistêmica individual têm correlatos sociais? Resultados que parecem apontar para essa direção são os de Zollman (2007) e Hahn et al (2018), que sugerem que a maior conectividade em redes de indivíduos está correlacionada com uma menor competência coletiva.

...

Por fim, penso que as noções de sanidade epistêmica, potencial alético e relevância epistêmica também podem ser utilizadas na elucidação de outras noções, como as noções de sabedoria teórica (Ryan, 2012), falsidades epistemicamente úteis (Elgin, 2019), etc. Porém, não tenho planos de lidar especificamente com esses pontos no futuro próximo.

OBJETIVOS

- Investigar as consequências de se tomar a sanidade epistêmica como um requisito da racionalidade para agentes finitos, em Epistemologia Individualista e Social.
 - Os sub-objetivos 1a, 1b, 1c, 2a, 2b, 3a, 3b, 4a e 4b foram discutidos acima.

MÉTODOS

Os objetivos listados acima são muito diferentes entre si e, conseqüentemente, demandam metodologias muito diferentes, que vão desde a tradicional análise conceitual (ex. objetivos 1b, 1c, 2a, 3a e 4b), a utilização de métodos formais (ex. 1a, 1b, 2b, 3a, 3b, 4a,

4b), a consideração de conceitos e resultados advindos de estudos empíricos (ex. objetivos 1a, 1b, 2a, 2b e 3b). Porém, o aspecto metodológico mais interessante e pervasivo neste projeto é a utilização de ferramentas da Epistemologia Computacional (ex. objetivos 1a, 1b, 1c, 2b, 3a, 3b, 4a e 4b). Em uma investigação típica de Epistemologia computacional, os epistemólogos projetam simulações computacionais, em que agentes artificiais (projetados para exibir a estratégia epistêmica a ser avaliada) atualizam suas crenças interagindo com ambientes (que podem conter outros agentes) gerados aleatoriamente a partir de alguns parâmetros fixos. Alguns exemplos de estudos em Epistemologia Computacional são Dantas (manuscrito), Trpin e Pellert (2019), Hahn et al (2018), Douven (2013), Olsson (2011) e Zollman (2007) em que Dantas (2020a) oferece uma revisão bibliográfica mais compreensiva. Neste projeto, pretendo investigar a noção de sanidade epistêmica implementando as medidas de potencial alético e relevância em simulações computacionais do Mundo de Wumpus (ver abaixo).

A necessidade do uso de ferramentas de Epistemologia Computacional se deve a algumas dificuldades em investigar como a racionalidade de agentes pode depender da estrutura dos ambientes (ex. objetivo 1b). Esta investigação requer medir o valor epistêmico de diferentes padrões de raciocínio em diferentes condições ambientais, mas esta medida dificilmente pode ser realizada com ferramentas experimentais e analíticas apenas porque é difícil acessar os valores de verdades das crenças de agentes reais e é difícil motivar a escolha de diferentes ambientes em argumentos analíticos. Esses problemas são evitados na Epistemologia Computacional, onde os epistemólogos têm acesso ao valor de verdade das crenças dos agentes artificiais e podem simular ambientes gerados aleatoriamente. O valor epistêmico de um padrão de raciocínio em num ambiente gerado aleatoriamente é um valor muito contingente, mas se o número de ambientes simulados for grande o suficiente, a média desses deve se aproximar do valor epistêmico esperado para aqueles ambientes.

Para a execução desse projeto, pretendo adaptar o modelo computacional que propus em meu doutorado e desenvolvi em Dantas (manuscrito). O modelo é baseado no Mundo de Wumpus, uma classe de problemas usada na IA para estudar o raciocínio incerto (Russell & Norvig, 2020, p. 210). O Mundo de Wumpus é uma caverna modelada como uma matriz de dimensões $s \times s$, em que (x, y) é a posição na x -ésima linha e y -ésima coluna da matriz e $(0, 0)$ é a entrada da caverna. Em uma sala aleatória (mas não na entrada), há um Wumpus, uma fera que come qualquer um que entre em sua sala. Todas as salas, exceto a entrada,

têm uma probabilidade independente de conter um poço, que também mata quem entrar. As salas diretamente adjacentes a uma sala que contém um poço [Wumpus] contém uma brisa [fedor]. O objetivo do agente é encontrar uma pilha de ouro localizada numa sala escolhida aleatoriamente (mas não a entrada) e retornar à entrada (“sair da caverna”). Na versão epistêmica do Mundo de Wumpus que desenvolvi, o agente também tem o objetivo de formar o conjunto de crenças mais compreensivo e acurado possível sobre as posições dos poços e do Wumpus. Além disso, há uma pequena probabilidade de ocorrerem brisas e fedores aleatórios, cuja ocorrência é independente da posição dos poços e do Wumpus.

Será necessário adaptar este modelo para cada um dos assuntos estudados. Por exemplo, para estudar o esquecimento, minha ideia seria repetir cada mapa gerado aleatoriamente em um número de tentativas, de modo que o agente possa ‘carregar’ informação obtida em tentativas anteriores para a próxima, a menos que tenha esquecido. Pretendo permitir que haja a possibilidade do mapa mudar parcialmente entre uma tentativa e outra (abrupta ou gradualmente)⁴. A investigação, então, poderia comparar a performance de agentes com diferentes padrões de esquecimento em ambientes com diferentes graus e tipo de mudança. Nos estudos de Epistemologia Social, pretendo utilizar o Laputa (www.laputasoftware.se/), um modelo apresentado em Olsson (2011) e utilizado em Hahn et al (2018). Nesse modelo, é possível escolher a estrutura da rede, o limiar de certeza para que cada agente afirme sua opinião através da rede, o grau de confiança que cada agente tem em relação a cada outro, a chance de agentes fazerem investigações individuais (e a confiabilidade dessas investigações), etc. Nesse caso também serão necessárias adaptações porque o modelo original só lida com uma crença por agente. Além disso, será necessário modificar a medida de valor epistêmico da utilizada em Laputa (V-value, cf. Goldman, 1999, p. 88) para a utilizada em nossa investigação (α -value, cf. justificção). Pretendo entrar em contato com Olsson para pedir autorização para modificar o código fonte de seu modelo.

JUSTIFICAÇÃO

A pesquisa sobre sanidade epistêmica é dificultada porque a maioria dos modelos formais utilizados em Epistemologia são inadequados para o estudo da evitação de entulho, esquecimento e suspensão de julgamento. A Epistemologia deveria estar minimamente preocupada com a racionalidade de raciocinadores finitos (afinal, somos raciocinadores

⁴ O aparecimento ou desaparecimento de um poço de uma dada sala da caverna seria uma mudança abrupta. A movimentação do Wumpus de uma sala para uma sala vizinha seria uma mudança gradual.

finitos), mas os epistemólogos frequentemente propõem modelos de racionalidade na forma de raciocinadores ideais (raciocinadores sem limitações cognitivas). Por exemplo, embora Leitgeb (2014, p. 137) reconheça que “em última análise, devemos nos preocupar com agentes do mundo real”, seu “raciocinador perfeitamente racional” é logicamente onisciente, o que é impossível para raciocinadores finitos⁵. A estratégia geral parece ser propor modelos de racionalidade enquanto raciocinadores ideais “os quais devemos nos esforçar para nos aproximar” (*idem*, nota de rodapé 3). Eu discordo dessa estratégia porque o que é racional para um raciocinador ideal pode não o ser para raciocinadores finitos⁶. Mesmo deixando essa questão de lado, a pesquisa sobre sanidade epistêmica ainda seria prejudicada porque os raciocinadores ideais que são usados como modelos de racionalidade geralmente não são capazes de evitar entulho, esquecer ou suspender.

O modelo bayesiano descreve o raciocinador ideal como uma função de graus de crença (*credence function*), que é consistente com os axiomas do cálculo de probabilidades (cf. Kolmogorov, 1950) e cujos valores são atualizados a partir de condicionalização bayesiana. O axioma da normalidade implica que o raciocinador deve estar certo (ter grau de crença máximo) sobre todas as verdades lógicas. Se o raciocinador aprende alguma evidência com certeza, então os axiomas da normalidade e da aditividade finita exigem que ele tenha certeza sobre todas as consequências lógicas da evidência (muitas das quais são ‘triviais’ no sentido acima). O raciocinador bayesiano é logicamente onisciente e, portanto, incapaz de evitar a desordem. Mas ele também é incapaz de esquecer. Se o raciocinador ideal atualiza seus graus de crença por condicionalização bayesiana, qualquer certeza que ele tenha antes de uma atualização permanecerá depois da atualização. Suponha que às 18h30 do dia 15/03/1989, um agente tenha certeza de que vai jantar espaguete naquela noite, mas em 15/03/1990 ele tenha esquecido completamente o que comeu um ano atrás (Talbot, 1991, p. 139). No exemplo, o agente perde certeza sobre uma proposição entre 15/03/1989 e 15/03/1990, o que é uma violação do modelo de racionalidade bayesiano⁷.

⁵ Um raciocinador logicamente onisciente acredita em todas as consequências lógicas de suas crenças, mas qualquer conjunto de crenças tem infinitas consequências lógicas (o que inclui as infinitas verdades lógicas). Porém, para qualquer noção minimamente realista de crença (em que crer requer da memória e capacidade computacional, cf. Dantas, 2019), um raciocinador com recursos cognitivos finitos não pode um número infinito de crenças. Além disso, há verdades lógicas com crescente complexidade e, para qualquer raciocinador finito, há verdades lógicas que ele não pode sequer entreter.

⁶ Por exemplo, por que deveríamos nos esforçar para aproximar a onisciência lógica? Qualquer tentativa de fazermos isso resultaria num tipo de paralisia cognitiva (nossos recursos cognitivos seriam desperdiçados na derivação de verdades e consequências lógicas), o que nos impediria de cumprir nossos objetivos.

⁷ Supondo que crença depende da memória. Titelbaum (2013) oferece um modelo (*certainty-loss framework*), que apresenta agentes próximos do bayesiano que são capazes de esquecer. Pretendemos avaliar esse modelo.

Seguindo Joyce (1998), alguns epistemólogos propuseram argumentos que usam princípios da Teoria da Decisão e medidas numéricas de valor epistêmico para motivar normas bayesianas, resultando na Teoria da Utilidade Epistêmica. Esses argumentos geralmente seguem três passos. O primeiro passo é definir o “conjunto ideal de crenças”: “se uma proposição é verdadeira numa situação, o grau de crença ideal para um agente naquela situação é o máximo, que é representado como 1. ... Se uma proposição é falsa, o grau de crença ideal nela é o mínimo, que é representada como 0” (Pettigrew, 2016, p. 3). Um agente com o conjunto ideal de crenças é incapaz de suspender, já que ele tem um grau de crença (extremo) sobre todas as proposições em sua agenda. O segundo passo é definir uma medida de (des)valor epistêmico (inacurácia), que é a distância entre os graus de crença de um agente e os graus de crença ideais correspondentes (medida usando um Brier score). Inacurácia não é uma medida de (des)valor epistêmico adequada para o estudo da suspensão de julgamento porque esta é uma maneira ‘barata’ de evitar inacurácia (ex. um cético tem zero inacurácia, cf. Carr, 2015)⁸. Retorno a esse ponto nos próximos parágrafos.

Como resultado da “estratégia geral”, o estudo da sanidade epistêmica e de seus principais contribuidores é muitas vezes deixado de lado na Epistemologia, especialmente naquela formalmente orientada. As limitações cognitivas dos raciocinadores finitos recebem uma consideração seletiva em Epistemologia. Como dito acima, os epistemólogos muitas vezes ignoram as limitações cognitivas dos raciocinadores finitos em seus modelos racionalidade (cf. Skipper & Bjerring, 2020, para uma exceção), mas tais limitações são citadas em algumas discussões, sobre a análise do conhecimento (Foley, 2019), a estrutura da justificação epistêmica (Huemer, 2019), o ceticismo moderado (Ryan, 2019), etc. Do mesmo modo, evitação de entulho, esquecimento e suspensão são investigados de maneira desigual na Epistemologia. A evitação da desordem é geralmente relegada a um tópico de interesse prático e, portanto, marginal à Epistemologia tradicional, que deve lidar com valor epistêmico. Os epistemólogos muitas vezes relegam o esquecimento e concentram suas investigações na lembrança bem-sucedida. Já a suspensão de julgamento ocupa uma posição central na Epistemologia desde seus primórdios, mas é discutida quase sempre no contexto do ceticismo (ex. Sextus Empiricus, 2000) e não na vida “normal” de agentes.

⁸ O terceiro passo é usar a medida de inacurácia e algum princípio da Teoria da Decisão em um argumento para alguma norma de racionalidade. O terceiro passo não é diretamente relevante para nossa discussão aqui.

Nos últimos anos, tenho avançado algumas propostas que podem ajudar no estudo da sanidade epistêmica. A investigação do valor epistêmico da evitação de entulho, esquecimento e suspensão e, em geral, a comparação entre conjuntos de crenças com diferentes números de crenças, requer não apenas uma medida global de imprecisão (quantidade de falsidade, f), mas também uma medida de abrangência (quantidade de verdade, t) e uma função que integre essas duas medidas em um único valor. Em Dantas (2021), proponho a função $\alpha(t,f)=(t-f)/(t+f+c)$, em que c é uma constante. Essa função é tal que cada nova crença verdadeira aumenta o valor α num ritmo menor ($\alpha(t+2, f) - \alpha(t+1, f) < \alpha(t+1, f) - \alpha(t, f)$), quando $t > f$. Essa característica pode ser interpretada como uma penalidade por número de crenças, que premia a parcimônia cognitiva. Em Dantas (2022), proponho que o potencial alético de um agente pode ser definido em termos de quanto valor α "extra" o agente espera alcançar, dado seu conjunto de crenças e padrão de inferência atuais. Esta seria a seguinte quantidade $\Delta\alpha = \alpha_1 - \alpha_0$, em que α_0 é o valor α do conjunto de crenças atual (inicial) do agente e α_1 é o valor α esperado para conjunto de crenças do agente após a investigação (ex. a média entre diversas tentativas na simulação). A relevância de uma crença estaria relacionada ao aumento de $\Delta\alpha$ que ela promove. A crença de que ϕ seria relevante para um agente sse $\Delta\alpha(B \cup \{\phi\}) > \Delta\alpha(B \setminus \{\phi\})$, em que $\Delta\alpha(B)$ é o potencial alético do agente se seu conjunto inicial de crenças fosse B .

Aspectos interdisciplinares

Nas últimas décadas, a fronteira entre a filosofia e as demais ciências tornou-se mais difusa. Os filósofos passaram a utilizar em suas pesquisas não apenas resultados científicos, mas também métodos científicos. Uma consequência desse fenômeno é a maior prevalência de métodos formais em Filosofia: não apenas a lógica clássica, mas também lógicas alternativas, cálculo de probabilidades, teoria da decisão/jogos, teoria da informação, teoria da computação, etc. Enquanto isso, parte considerável do ensino e pesquisa em filosofia no Brasil se restringe à discussão dos filósofos a partir (quase que exclusivamente) da leitura de textos clássicos. Isso é uma situação ruim porque a cooperação interdisciplinar e a disciplina formal tem gerado bons resultados em Filosofia.

Os aspectos interdisciplinares deste projeto são óbvios, pois ele se baseia fortemente em teorias e resultados empíricos. Por exemplo, dados empíricos relevantes sobre a memória incluem aqueles sobre como o conteúdo da memória declarativa é estruturado (por exemplo, dados sobre erros de memória, cf. Dantas, 2020b), esquecimento incidental e

motivado, a relação entre esquecimento e raciocínio (por exemplo, heurísticas de raciocínio que exploram o esquecimento) e as limitações cognitivas dos seres humanos (por exemplo, limitações informativas na memória de trabalho). Além disso, o projeto também se baseia fortemente em métodos formais e computacionais. Os métodos computacionais abrangem técnicas-padrão em IA, como simulação baseada em agente, algoritmos de *pathfinding*, algoritmos de inferência a partir de redes bayesianas causais, mas também técnicas da teoria da computação, como o benchmark e a análise assintótica.

BIBLIOGRAFIA

- ANDERSON, J. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14(3), pp. 471-485.
- ANDONOVSKI, N. (2020). Memory as triage: Facing up to the hard question of memory. *Review of Philosophy and Psychology*, 12, pp. 227–256.
- CARR, J. (2015). Epistemic expansions. *Res Philosophica*, 92 (2), pp. 217–236.
- CHERNIAK, C. (1986). *Minimal Rationality*. MIT Press.
- DANTAS, D. (2022). Epistemic sanity or why you shouldn't be opinionated or skeptical. *Episteme*, pp. 1-20. <https://www.doi.org/10.1017/epi.2022.43>
- _____. (2021). How to (blind)spot the truth: An investigation on actual epistemic value. *Erkenntnis*. <https://doi.org/10.1007/s10670-021-00377-x>
- _____. (2020a). Epistemologia computacional: Uma provocação. *Perspectiva Filosófica*, 46(2), 189–221. <https://doi.org/10.51359/2357-9986.2019.248089>
- _____. (2020b). Two informational theories of memory: A case from memory-conjunction errors. *Disputatio*, 12(59), pp. 395–431. <https://doi.org/10.2478/disp-2020-0019>
- _____. (2019). Logical omnipotence and two notions of implicit belief. In T. Rodrigues (Org). *Epistemologia analítica: Debates contemporâneos*, 1st ed. (pp. 29–46). Editora Fi. <https://www.editorafi.org/665epistemologia>
- _____. (2017). No rationality through brute-force. *Philosophy South (Filosofia Unisinos)*, 18(3), pp. 195-200. <https://doi.org/10.4013/fsu.2017.183.11>
- _____. (manuscrito). When the (Bayesian) ideal is not ideal. tiny.cc/dantas_ideal
- DE BRIGARD, F. (2013). Is memory for remembering?: Recollection as a form of episodic hypothetical thinking. *Synthese*, 191(2), pp. 155–185.

- DOUVEN, I. (2013). Inference to the best explanation, Dutch books, and inaccuracy minimisation. *The Philosophical Quarterly*, 63(252), pp. 428–444.
- ELGIN, C. (2019). Epistemically useful falsehoods. In B. Fitelson, R. Borges, & C. Braden (Eds.), *Themes from Klein* (pp. 25–38). Springer.
- FOLEY, R. (2019). Accurate enough, comprehensive enough, and reasonable enough belief. In B. Fitelson, R. Borges, & C. Braden (Eds.), *Themes from Klein* (pp. 57–70). Springer.
- FRIEDMAN, J. (2020). The Epistemic and the Zetetic. *The Philosophical Review*, 129(4), pp. 501–536.
- _____. (2018). Junk beliefs and interest-driven Epistemology. *Philosophy and Phenomenological Research*, 97(3), pp. 568–583.
- _____. (2015). Why suspend judging? *Noûs*, 51(2), pp. 302–326.
- _____. (2013). Suspended judgment. *Philosophical Studies*, 162(2), pp. 165–181.
- GAO, J. (2021). Self-deception and pragmatic encroachment: A dilemma for epistemic rationality. *Ratio*, 34, pp. 20–32.
- GIGERENZER, G (2021). Axiomatic rationality and ecological rationality. *Synthese*, 198, pp. 3547–3564.
- GOLDMAN, A. (1999a). Internalism exposed. *The Journal of Philosophy*, 6, pp. 271–293.
- _____. (1999b). *Knowledge in a social world*. Clarendon Press.
- GRIFFITHS, T., LIEDER, F. & GOODMAN, N (2015), Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7, pp. 217–229.
- HAHN, U., HANSEN, J., OLSSON, E. (2018) Truth-tracking performance of social networks: how connectivity and clustering can make groups less competent. *Synthese*, 197(4), pp. 1511–1541.
- HARMAN, G. (1986). *Change in view: Principles of reasoned revision*. MIT Press.
- HUEMER, M. (2019). Finite minds. In B. Fitelson, R. Borges, & C. Braden (Eds.), *Themes from Klein* (pp. 171–187). Springer.
- KLEIN, S., ROBERTSON, T., & DELTON, A. (2009). Facing the future: Memory as an evolved system for planning future acts. *Memory & Cognition*, 38(1), pp. 13–22.
- KOLMOGOROV, A. (1950). *Foundations of probability*. Chelsea Publishing Company.

- LEITGEB, H. (2014). The stability theory of belief. *The Philosophical Review*, 123(2), pp. 131–171.
- JOYCE, J. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4), 575–603.
- LORD, E. (2020). Suspension of judgment, rationality's competition, and the reach of the epistemic. In S. Schmidt & G. Ernst (Eds.), *The Ethics of Belief and beyond: Understanding mental normativity* (pp. 126–145). Routledge.
- MARR, D. (1982). *Vision: A Computational Approach*. Freeman & Co.
- MICHAELIAN, K. (2011). The Epistemology of forgetting. *Erkenntnis*, 74(3), pp. 399–424.
- OAKSFORD, M., & CHATER, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
- OLSSON, E. (2011). A simulation approach to veritistic social Epistemology. *Episteme*, 8(2), pp. 127–143.
- PETTIGREW, R. (2019). Veritism, epistemic risk, and the swamping problem. *Australasian Journal of Philosophy*, 97(4), pp. 761-774.
- _____. (2016). *Accuracy, chance, and the laws of credence*. Oxford University Press.
- POROT, N & MANDELBAUM, E. (2021). The science of belief: A progress report. *Wiley Interdiscip Rev Cogn Sci.*, 12, e1539.
- RUSSELL, S., & NORVIG, P. (2020). *Artificial intelligence: A modern approach* (4th Edition, US Edition). Pearson.
- RYAN, S. (2019). Epistemic humility, defeat, and a defense of moderate skepticism. In B. Fitelson, R. Borges, & C. Braden (Eds.), *Themes from Klein* (pp. 129–143). Springer.
- _____. (2012). Wisdom, knowledge and rationality. *Acta Analytica*, 27(2), pp. 99–112.
- SEXTUS EMPIRICUS (2000). *Outlines of scepticism* (J. Annas & J. Barnes, Eds.; 2nd ed.). Cambridge University Press.
- SHANAHAN, Murray, The frame problem. *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (Ed.),
URL=<<https://plato.stanford.edu/archives/spr2016/entries/frame-problem/>>.
- SKIPPER, M., & BJERRING, J. (2020). Bayesianism for non-ideal agents. *Erkenntnis*, 87, pp. 93–115

- SINGER, D, BRAMSON, A., GRIM, P. et al (2021). Don't forget forgetting: the social epistemic importance of how we forget. *Synthese*, 198, pp. 5373–5394.
- STURGEON, S. (2010). Confidence and coarse-grained attitudes. In T. Gendler & J. Hawthorne (Eds.). *Oxford studies in Epistemology*, volume 3 (pp. 126–149). Oxford University Press.
- _____. (2008). Reason and the grain of belief. *Noûs*, 42(1), 139–165.
- TALBOTT, W. (1991). Two principles of Bayesian Epistemology. *Philosophical Studies*, 62(2), 135–150.
- TITELBAUM, M. (2013). *Quitting certainties: a Bayesian framework modeling degrees of belief*. Oxford University Press.
- TRPIN, B. & PELLERT, M. (2019). Inference to the best explanation in uncertain evidential situations. *British Journal for the Philosophy of Science*, 70(4), pp. 977–1001.
- VAN ROOIJ, I. (2008). The tractable cognition thesis. *Cognitive Science*, 32(6), pp. 939-984.
- ZOLLMAN, K. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5), pp. 574-587.