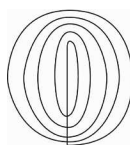


# FIABILISMO

EDIÇÃO DE 2018 do

## COMPÊNDIO EM LINHA DE PROBLEMAS DE FILOSOFIA ANALÍTICA

2012-2015 FCT Project PTDC/FIL-FIL/121209/2010



Editado por  
Ricardo Santos e Pedro Galvão

ISBN: 978-989-8553-22-5

Compêndio em Linha de Problemas de Filosofia Analítica  
Copyright © 2018 do editor  
Centro de Filosofia da Universidade de Lisboa  
Alameda da Universidade, Campo Grande, 1600-214 Lisboa

Fiabilismo  
Copyright © 2018 do autor  
Ernesto Perini-Santos

DOI: <https://doi.org/10.51427/cfi.2021.0040>

Todos os direitos reservados

**Resumo**

A tese central do fiabilismo é que uma crença verdadeira é conhecimento apenas se foi produzida por um mecanismo que tende a gerar crenças verdadeiras. Como tanto o processo que gera uma dada crença quanto sua propensão a produzir crenças verdadeiras podem não ser apreendidos pelo sujeito a quem se atribui o conhecimento, o fiabilismo é uma teoria externista. A principal fonte de críticas ao fiabilismo reside precisamente na desvinculação entre o que torna crenças meramente verdadeiras conhecimento e a perspectiva do sujeito. As diferentes versões do fiabilismo são diferentes respostas a esta crítica, quer pela aproximação com as ciências cognitivas e com o projeto naturalista, quer pela proposição de uma epistemologia da virtude.

**Palavras-chave**

Fiabilismo, epistemologia, externismo, epistemologia da virtude, naturalismo

**Abstract**

The central claim of reliabilism is that a true belief counts as knowledge only if it has been generated by a mechanism that tends to produce true beliefs. Reliabilism is an externalist theory, since both the process generating any given belief and its tendency to lead to true beliefs may not be apprehended by the subject to which knowledge is attributed. The central criticism to reliabilism is precisely that what distinguishes mere true belief from knowledge may be external to the subject's perspective. The responses to this criticism lead to different versions of reliabilism, either moving towards cognitive science and a naturalist project or proposing a virtue epistemology.

**Keywords**

Reliabilism, epistemology, externalism, virtue epistemology, naturalism

# Fiabilismo

DOI: <https://doi.org/10.51427/cfi.2021.0040>

## 1 Introdução

A tese central do fiabilismo, formulada inicialmente por Alvin Goldman, pode ser apresentada do seguinte modo: uma crença verdadeira é conhecimento apenas se foi produzida por um mecanismo que tende a gerar crenças verdadeiras.<sup>1</sup> Como uma crença pode ser obtida a partir de uma base não doxástica ou a partir de outra crença, a exigência fiabilista porta tanto sobre processos de obtenção de crenças a partir de uma base não doxástica quanto sobre crenças que derivam de outras crenças; neste último caso, o processo deve manter a fiabilidade da crença ou das crenças originais.

O fiabilismo pode inicialmente ser formulado sem referência ao processo de obtenção de crenças, através de condicionais contrafactuais:

1. Uma crença  $p$  de um sujeito  $S$  é conhecimento se, (i) se  $p$  não for verdadeira,  $S$  provavelmente não acredita que  $p$  e (ii) se  $p$  for verdadeira,  $S$  provavelmente acredita que  $p$ .

A crença  $p$  de  $S$  rastreia a verdade.<sup>2</sup> A proposta (1) deve ser revista. Inicialmente, devemos evitar que as duas condições sejam avaliadas relativamente a métodos diferentes. Considere o seguinte exemplo de Robert Nozick:

Uma avó vê que seu neto está bem de saúde quando ele a visita. No entanto, se ele estivesse doente ou morto, para poupá-la de dissabores, as outras pessoas diriam-lhe que ele está bem de saúde. Isto no entanto não quer dizer que ela não saiba que ele está bem de saúde (ou pelo menos que não está acamado) quando o vê. Devemos claramente reformular nossas condições para levar explicitamente em conta as maneiras

<sup>1</sup> “O estatuto de justificabilidade de uma crença é uma função da fiabilidade do processo ou dos processos que a causam, onde (numa primeira aproximação) a fiabilidade consiste na tendência de um processo para produzir crenças que são verdadeiras ao invés de falsas.” Goldman: 2011 (Goldman 1992: 113). Nas citações de Goldman 2011, ‘confiável’ foi substituído por ‘fiável’.

<sup>2</sup> “Conhecer é ter uma crença que rastreia a verdade. O conhecimento é uma maneira particular de se estar conectado com o mundo, através de uma conexão real específica factual com o mundo : rastreando-o.” Nozick 1981: 178.

e os métodos para se chegar a uma crença.<sup>3</sup>

Neste caso, um método satisfaz (ii), mas a condição (i) é violada por outro método. Parece no entanto razoável dizer que, ao ver seu neto, ela sabe que ele está bem de saúde, contrariamente ao que prevê (1). Além disto, como observa Colin McGinn, a definição de Nozick carece de generalidade: ela diz respeito apenas a uma proposição e o conhecimento resulta da “propensão de uma pessoa a acreditar a verdade em relação a um conjunto determinado de proposições relevantes”. A generalização requer um método que rastreie a verdade para um dado conjunto de proposições.<sup>4</sup>

A reformulação nos aproxima da proposta de Goldman :

2. Uma crença  $p$  de um sujeito  $S$  é conhecimento se foi obtida por método  $M$  tal que, através de  $M$ , (i) se  $p$  não for verdadeira,  $S$  provavelmente não acredita que  $p$  e (ii) se  $p$  for verdadeira,  $S$  provavelmente não acredita que  $p$ .

(2) é uma condição necessária, mas não suficiente para o conhecimento. Acrescentando a condição de verdade e a exigência do sujeito tomar por verdadeiro aquilo que sabe, chegamos à definição tripartite do conhecimento:

(D) $S$ sabe que $p$ se e somente se	(i) $p$ é verdadeira,
	(ii) $S$ acredita que $p$ e
	(iii) $S$ está justificado em acreditar que $p$ porque $p$ foi obtida por um processo fiável de obtenção de crenças.

A tese fiabilista<sup>5</sup> assim exposta é uma teoria acerca da condição que deve ser acrescentada a uma crença verdadeira para que ela seja dita

<sup>3</sup> Nozick 1981: 179.

<sup>4</sup> McGinn 1999: 17.

<sup>5</sup> Para a exposição da tese fiabilista utilizarei ‘processo’ e ‘mecanismo’ ao invés de ‘método’. Embora estes três termos não sejam equivalentes, sua diferença não será importante para a apresentação geral do fiabilismo.

conhecimento.<sup>6</sup> É possível no entanto defender uma teoria fiabilista sem se adotar a análise tripartite do conhecimento, como propõe notadamente Tim Williamson.

A fiabilidade do processo de produção de crenças depende do seu contexto de exercício e é relativo a um determinado domínio proposicional. A articulação do fiabilismo irá precisamente definir melhor o que é um processo fiável nestas duas dimensões. A principal fonte de críticas ao fiabilismo reside na desvinculação entre o que torna crenças meramente verdadeiras conhecimento e a perspectiva do sujeito sobre sua própria situação epistêmica. Além disto, o fiabilismo enfrenta o desafio de delimitar o tipo sob o qual cai o processo particular de produção de crenças a ser avaliado. As respostas a estas críticas levam a diferentes versões do fiabilismo, a epistemologia da virtude, por um lado, e o caminho da naturalização da epistemologia, por outro.

## 2 Articulando o Fiabilismo

O fiabilismo foi proposto inicialmente por Goldman como uma resposta ao desafio de Gettier à definição tripartite do conhecimento.<sup>7</sup> Uma versão do fiabilismo é uma teoria acerca da terceira condição:

(D-iii) *S* está justificado em acreditar que *p* porque *p* foi obtida por um processo fiável de obtenção de crenças.

Um gato preto está agora andando na sala do apartamento de Ana. Arnaldo acredita que há um gato preto na sala do apartamento de Ana porque a borra de café que restou na sua xícara lembra um gato. Embora sua crença seja verdadeira, ele não sabe que há um gato preto andando na sala do apartamento de Ana, porque a condição (D-iii) não está preenchida. Ana adquire esta mesma crença ao entrar em

<sup>6</sup> Sosa diz assim, no início de recente apresentação de sua versão do fiabilismo: “Deste modo, nossa questão é a seguinte: que condição uma crença deve satisfazer, além da verdade, para ser conhecimento”. Sosa 2011: 2. Mais sobre Sosa abaixo.

<sup>7</sup> Goldman “What is Justified Belief?” (1979), republicado em Goldman 1992 e traduzido em Goldman 2011. Sobre a definição tripartite e o desafio de Gettier, ver Rodrigues 2013

seu apartamento e ver que há um gato preto na sala. Ela sabe há um gato preto na sala de seu apartamento. Segundo o fiabilismo, o contraste entre as duas situações resulta do fato de a crença de Ana ter sido produzida por um processo que tende a gerar crenças verdadeiras.

O fiabilismo é uma teoria genética da justificação. Uma consequência do foco na gênese da crença é a desvinculação entre o que torna o sujeito justificado em tomar algo por verdadeiro e aquilo que é lhe acessível conscientemente: nós não temos em geral acesso introspectivo a nossos mecanismos cognitivos. Além disto, a propensão a gerar crenças verdadeiras de um dado processo cognitivo é independente da sua apreensão pelo sujeito. Estas duas condições não dependem tampouco apenas de estados mentais do sujeito. O internismo em epistemologia pode requerer que o sujeito tenha acesso consciente ao que torna sua crença justificada<sup>8</sup> ou demandar apenas que a justificação seja superveniente sobre estados mentais do sujeito.<sup>9</sup> O fiabilismo nega estas duas condições e é assim uma teoria externista da justificação. Não é necessário que Ana tenha refletido, ou mesmo seja capaz de refletir sobre o modo como a percepção tende a levar a crenças verdadeiras para estar justificada em acreditar que um gato anda na sala do seu apartamento em virtude da fiabilidade de sua visão. Além disto, o fato de um determinado processo, por exemplo, a visão, tender a levar a crenças verdadeiras não depende apenas do que ocorre no sujeito que adquire uma crença a partir de sua experiência visual.

Mas será a visão sempre um mecanismo fiável de obtenção de crenças? Inicialmente, a visão é uma base fiável para a obtenção de crenças de um determinado tipo. Agora, você acredita ter um texto escrito a sua frente, uma crença verdadeira que resulta de sua experiência visual. No entanto, se você devesse julgar se esta frase foi escrita numa quinta-feira à tarde ou num dia ensolarado, sua experiência visual presente não seria de muita utilidade. Além disto, a experiência visual é uma base fiável em determinados contextos, mas não em todo contexto: se este texto estivesse numa sala escura ou a cinco metros de distância, sua experiência visual não lhe seria de

<sup>8</sup> E.g., Chisholm 1989.

<sup>9</sup> E.g., Conee e Feldman 2004; Wedgwood 2002.

grande utilidade. Sua experiência visual não seria tampouco útil se tivesse tomado um chá que lhe turvasse a visão. Há uma explicação para estas limitações: a visão é sensível a um determinado tipo de informação do ambiente e sua boa performance depende de algumas condições internas e externas ao sujeito. Ao explicar a justificação deste modo, o fiabilismo fornece uma base não epistêmica a nossos julgamentos epistêmicos: a propensão de um processo de produção de crenças de levar a crenças verdadeiras em determinadas circunstâncias pode ser explicada sem recurso a um termo avaliativo.<sup>10</sup>

Podemos modificar a definição inicial do seguinte modo:

(D-iii') *S* está justificado em acreditar que *p* porque sua crença foi produzida por um processo de obtenção de crenças fiável relativamente a um contexto de tipo *C* e a um campo proposicional *F* no qual se encontra *p*.

Como já vimos, McGinn critica o caráter limitado da proposta inicial de Nozick. Segundo McGinn, (1) define apenas a fiabilidade local,<sup>11</sup> em oposição à fiabilidade global, relativa a uma determinada classe de proposições:

(...) dizer que uma pessoa *S* é globalmente fiável em relação a um conjunto de proposições é dizer que *S* pode *discriminar* a verdade da falsidade dentro deste conjunto de proposições (...)<sup>12</sup>

Esperamos que a capacidade de rastrear a verdade ilustrada em (1) resulte do exercício de uma capacidade mais geral, que rastreie também a verdade de um conjunto de proposições similares. Sem esta capacidade

<sup>10</sup> “O termo ‘justificado’, presumo, é usado para atribuir valor, é um termo apreciativo. Em qualquer definição correta ou em qualquer sinônimo seu pareceriam termos avaliativos. Presumo que tais definições ou sinônimos poderiam ser oferecidos, mas não estou interessado neles. Quero um conjunto de condições *substanciais* que especifiquem quando uma crença é justificada. (...) quero uma teoria da crença justificada que especifique em termos não epistêmicos quando uma crença é justificada.” Goldman 2011 (1992: 105)

<sup>11</sup> “Diremos que (...) a teoria do rastreamento de Nozick é uma análise *local* do conhecimento – ela localiza as condições do conhecimento numa relação entre aquele que conhece e uma única proposição.” McGinn 1999: 16.

<sup>12</sup> McGinn 1999: 17.

mais ampla, diz Edward Craig, a satisfação de 1 é pura sorte.<sup>13</sup>

A oposição entre o relativismo global e local pode ser feita de maneira diferente daquela proposta por McGinn. Inicialmente, como já vimos no caso da visão, a fiabilidade é também relativa ao contexto de exercício de uma dada capacidade, e não apenas a um tipo determinado de proposição. Além disto, a oposição mesma entre fiabilidade local e fiabilidade global pode ser traçada de diferentes maneiras. David Henderson e Terence Horgan, por exemplo, identificam a fiabilidade local à fiabilidade relativa a um ambiente específico ocupado pelo agente cognitivo e a fiabilidade global a uma capacidade relativa a um conjunto mais amplo de situações.<sup>14</sup> Contrariamente à proposta de McGinn, a fiabilidade local não se limita a uma única proposição, mas a um ambiente particular, com características atípicas relevantes para o julgamento da fiabilidade de um dado mecanismo cognitivo que não se estende a outros ambientes. Esta é a maneira mais usual de se contrastar a fiabilidade local e a fiabilidade global, opondo ambientes mais restritos e ambientes mais amplos.

Ana sabe em geral reconhecer gatos. O seu bairro está no entanto infestado de gatos-robô, que ela não sabe distinguir de gatos reais. Todos os gatos do bairro são de cor cinza, enquanto os robôs são de outra cor. Ela vê um gato cinza no corredor e pensa que se trata de um gato. Arnaldo, por outro lado, não é em geral muito bom para reconhecer gatos. Ele acredita em particular que gatos podem ter apenas a cor cinza. Ele também acredita o gato cinza a sua frente é um gato. A crença de Ana foi produzida por um mecanismo de reconhecimento de gatos globalmente fiável, mas localmente não fiável. Arnaldo, inversamente, tem um mecanismo localmente fiável, mas globalmente ineficiente de reconhecimento de gatos. A mera fiabilidade não é o bastante para distinguir a robustez das crenças de Ana em relação a gatos, em oposição à fragilidade das crenças de Arnaldo. É um acaso que as crenças de Arnaldo, mas não as de Ana, sejam fiáveis neste ambiente particular. A fiabilidade do mecanismo de formação de crenças de Ana sobre gatos é mais robusta, vale num conjunto maior de situações e, por esta razão, ela está justificada em

<sup>13</sup> Craig 1990: 58-59.

<sup>14</sup> Horgan e Henderson 2011: 68.



suas crenças, mesmo se ele não é localmente fiável.<sup>15</sup>

Não há fronteiras fixas entre o que conta como local e o que conta como global – na situação dos gatos-robô, o caráter local deriva da atipicidade do ambiente que leva a estes resultados, quando comparado com o conjunto de ambientes nos quais Ana e Arnaldo costumam viver. Tanto a oposição global/local, quando a oposição típico/atípico são vagas, mas a noção intuitiva de fiabilidade também parece ser vaga.<sup>16</sup> Uma crítica ao fiabilismo chamada de problema da generalidade explora precisamente este ponto, como veremos abaixo.

Uma escolha possível neste ponto é a expansão máxima de ambientes relevantes para a avaliação de capacidade cognitivas, i.e., capacidade cognitivas são avaliadas em mundos. No chamado fiabilismo modal, a fiabilidade de um dado processo de obtenção de crenças é comparado através de mundos possíveis. Há aqui um efeito de argumentos céticos que exploram o estatuto de nossas crenças em mundos epistemicamente hostis. Imaginemos um mundo  $m'$  no qual um gênio maligno produza num sujeito  $S$  experiências perceptuais indistinguíveis das experiências no mundo efetivo, mas que a nada correspondem em  $m'$ . Neste caso, as crenças de  $S$  resultantes de experiências perceptuais não tendem a ser verdadeiras, mas falsas, e a experiência perceptual não será um processo fiável de produção de crenças. Segundo o veredito fiabilista, crenças assim obtidas não estão justificadas em  $m'$ , porque são globalmente não fiáveis – voltaremos a esta avaliação abaixo. A fiabilidade de um dado processo será assim relativa a mundos.

Há uma outra razão para a relativização a situações amplas: explicações para o funcionamento de capacidades exercidas na aquisição do conhecimento são relativas ao tipo de ambiente no qual vive e evoluiu uma determinada espécie ou uma determinada capacidade. Por um lado, não há razão para se esperar que a fiabilidade de processos de obtenção de crenças se estenda a situações muito diferentes daquelas nas quais evoluiu um dado organismo ou uma dada capacidade cognitiva. Por outro lado, os mecanismos cognitivos eles mesmos são sensíveis a determinados parâmetros ambientais, mas não a todo

<sup>15</sup> Esta é uma versão modificada de um exemplo de Horgan e Henderson: 2011: 63-73.

<sup>16</sup> Alston 1995: 10.

ambiente possível, nem a todo aspecto de um dado ambiente. Retomando a distinção de Ernst Mayr entre os dois tipos de perguntas em explicações biológicas, sobre o como e sobre o por que de um dado traço, tanto a pergunta sobre como um mecanismo cognitivo funciona quanto a pergunta evolutiva acerca de por que ele existe são relativas a ambientes.<sup>17</sup> Como o fiabilismo busca uma base não epistêmica para conceitos epistêmicos, a explicação do como e do por que de nossas capacidades cognitivas ocupa um lugar central na teoria. Esta explicação é restrita a mundos similares ao mundo de exercício de uma capacidade determinada.

A relativização a contextos permite uma precisão conceitual importante acerca da natureza mesma da fiabilidade. Um mecanismo que leva à crença  $p$  deve excluir alternativas a  $p$ , i.e., situações nas quais  $p$  é falsa. Nenhum mecanismo é no entanto capaz de excluir todas tais alternativas; este é precisamente o ponto explorado por argumentos céticos. A consequência que tira o fiabilista não é, como quer o cético, de abandonar a fiabilidade de nossos mecanismos cognitivos em geral, mas impor uma exigência mais fraca: um mecanismo de produção de crenças é fiável num dado contexto  $C$  apenas se for capaz de excluir as alternativas relevantes em  $C$ . Como definir a classe de alternativas relevantes?

Fred Dretske aponta uma solução: crenças são adquiridas contra o pano de fundo de pressuposições que delimitam o espaço de alternativas relevantes.<sup>18</sup> Um célebre exemplo de Dretske ilustra esta tese:

Você leva seu filho ao zoológico, vê várias zebras e, quando se filho pergunta o que são, você lhe diz que são zebras. Você sabe se são zebras? Bem, a maior parte de nós não hesitaria em dizer que sim. Nós conhecemos a aparência de zebras e, além disto, estamos no zoológico municipal e os animais estão numa jaula onde está claramente escrito 'Zebra'. Mas que algo seja uma zebra implica que não seja uma mula, que não seja, em particular, uma mula habilmente disfarçada como zebra pelas autoridades do zoológico. Você sabe se estes animais não são

<sup>17</sup> "O biólogo funcionalista (*functional biologist*) está essencialmente preocupado com a operação e a interação de elementos estruturais, das moléculas até órgãos e indivíduos como um todo. Sua questão sempre é «Como?» (...) O biólogo evolucionário difere em seus métodos e nos problemas nos quais está interessado. Sua pergunta básica é «Por que?»". Mayr 1961: 1502. Ver a distinção entre causa estruturante e causa desencadeadora em Dretske 1988: 42-44.

<sup>18</sup> Dretske 1970.

mulas habilmente disfarçadas de zebras pelas autoridades do zoológico?<sup>19</sup>

Um cético pode argumentar uma resposta positiva exige a exclusão da alternativa de os animais na jaula serem mulas disfarçadas de zebra – o que você certamente não fez. Se a conclusão cética é que você não sabe se de fato estes animais são zebras, a resposta fiabilista é que você deve excluir esta alternativa apenas se ela for relevante. Como suas pressuposições acerca do comportamento de administradores de zoológicos não tornam relevante esta alternativa, ela não deve ser excluída. Pressuposições não explicam toda delimitação de circunstâncias nas quais mecanismos cognitivos são ou não fiáveis. As limitações devidas a traços de nosso aparato cognitivo não dependem de atitudes proposicionais do sujeito. Ana não é capaz de saber as diferenças de temperatura no seu ambiente pela experiência visual, como o faria, por exemplo, se tivesse uma visão análoga a uma câmera térmica. Esta limitação não depende de algo que Ana acredite, mas de um traço do aparato visual humano, a saber, a limitação do espectro de luz visível por nós. Ela não faz julgamentos sobre a temperatura do ambiente a partir da visão, e esta limitação não se segue do que ela pressupõe – é indiferente que ela o pressuponha ou não que tem esta limitação –, mas de um traço do aparato visual humano.

Podemos reformular a condição (iii) do seguinte modo:

(D-iii'') *S* está justificado em acreditar que *p* porque sua crença foi produzida por um processo de obtenção de crenças *M* fiável relativamente a um contexto de tipo *C*, porque *M* é capaz de excluir alternativas a *p* relevantes em contextos de tipo *C*, e a um campo proposicional *F* no qual se encontra *p*.

Mecanismos cognitivos podem ser mais ou menos fiáveis em diferentes ambientes. Em todo ambiente, no entanto, mecanismos cognitivos são falíveis. Para toda crença *p*, tudo o que impeça *p* de ser conhecimento, por exemplo, por tornar *p* falsa, ou por minar a fiabilidade do processo que levou à crença que *p*, é dito ser um “derrotador” (*defeater*) do conhecimento que *p*. Para que uma crença *p* seja justificada em *C* não basta que ela seja o resultado de um mecanismo fiável, é preciso que não existam derrotadores do conhecimento que

<sup>19</sup> Dretske 1970: 1016.

$p$  em  $C$ . Assim, se Ana tiver tomado um chá que faz com que metade de suas experiências visuais sejam alucinações, sua experiência de visual de ver um gato não será mais um indicador fiável da presença de um gato. O fato de ela ter tomado um chá mina a fiabilidade da sua experiência visual. Ela não saberá tampouco que há um gato a sua frente se estiver vendo um robô. A crença de Ana está justificada, mesmo se houver derrotadores? Estas são alternativas relevantes que deveriam ter sido excluídas por Ana?

Goldman, no seu artigo de 1979, inclui uma cláusula contemplando a possibilidade de derrotadores:

Se a crença de  $S$  em  $p$  no momento  $t$  for o resultado de um processo cognitivo fiável, e não houver um processo fiável ou condicionalmente fiável disponível a  $S$  que, caso tivesse sido usado por  $S$ , além do processo realmente usado, levaria  $S$  a não acreditar em  $p$  em  $t$ , então a crença de  $S$  em  $p$  no momento  $t$  está justificada.<sup>20</sup>

A justificação de uma dada crença não depende apenas de ela ter sido produzida por um mecanismo fiável, mas não deve haver derrotadores que poderiam ter sido detectados por algum mecanismo cognitivo disponível então para o sujeito. As descrições dos casos acima estão sub-determinadas: Ana sabe que tomou um chá alucinógeno? Ela tem a capacidade de distinguir gatos de gatos-robô? Para nossa exposição, vamos considerar que esta condição está incluída na consideração de alternativas relevantes a uma dada situação.

Em (D-iii’),  $C$  compreende tanto situações globais máximas, i.e., mundos, quanto circunstâncias mais locais de exercício de uma capacidade cognitiva – no nosso exemplo inicial, o contexto de avaliação da crença de Ana compreende tanto o mundo efetivo, no qual a visão humana é em geral fiável, quanto o fato de não haver gatos-robô no seu apartamento. O segundo exemplo muda o contexto e altera a fiabilidade local de suas crenças. A exclusão de alternativas relevantes, descrita por condicionais contrafactuais, pode ser apresentada de maneira mais fina. Em *Discrimination and Perceptual Knowledge* (1976), Goldman explica a fiabilidade de processos cognitivos do seguinte modo:

Que tipos de processos ou mecanismos causais devem produzir uma crença para que ela conte como conhecimento? Eles devem ser me-

<sup>20</sup> Goldman 2011 (1992: 123).

canismos “fiáveis”, em um sentido apropriado do termo. Em termos gerais, um mecanismo ou processo é fiável se ele não apenas produz crenças verdadeiras em situações efetivas, mas produzir crenças verdadeiras, ou pelo menos inibir crenças falsas, em situações contrafactuais relevantes.<sup>21</sup>

Um mecanismo fiável deve passar um teste contrafactual expresso por dois condicionais subjuntivos:

(TC – *Segurança*): a crença de *S* que *p* é segura se, e apenas se, se *S* acreditasse em *p*, provavelmente *p* seria verdadeiro.

(TC – *Sensibilidade*): a crença de *S* que *p* é sensível se, e apenas se, se *p* fosse falso, *S* provavelmente não acreditaria em *p*.<sup>22</sup>

Um mecanismo de produção de crenças é fiável se ele produzir crenças sensíveis e seguras – por extensão, este mecanismo ele mesmo será sensível e seguro. As duas condições não são equivalentes, condicionais subjuntivos não se contrapõem. Segundo um cenário usual, minha crença que eu não sou um cérebro numa cuba é segura : nos mundos possíveis similares ao mundo efetivo, minha crença que eu não sou um cérebro numa cuba é verdadeira. No entanto, seu eu fosse um cérebro numa cuba estimulado por um cientista que me faz crer que estou agora lendo um texto numa tela de computador, eu provavelmente ainda acreditaria que não sou um cérebro numa cuba, minha crença portanto não é sensível. Estes condicionais subjuntivos ilustram a restrição da validade de mecanismos cognitivos a situações similares à situação de avaliação de uma crença ou ainda, no vocabulário dos mundos possíveis, em mundos possíveis próximos do mundo de avaliação.<sup>23</sup> Um mecanismo fiável de produção de crenças não é um mecanismo infalível, mas um mecanismo seguro e sensível.

O fiabilismo pode se apresentar não como uma teoria da justificação, mas diretamente como uma teoria do conhecimento. Os artigos de Goldman anteriores à proposta do fiabilismo, “A causal theory of knowing” (1967) e “Discrimination and perceptual knowledge”

<sup>21</sup> Goldman 1992: 85.

<sup>22</sup> Sosa 2013a: 35 (2007: 25). Esta tese origina-se na teoria de Nozick que uma crença é conhecimento se resultar de um mecanismo que rastreie a verdade; Nozick 1981: 172-178.

<sup>23</sup> Nozick 1981: 176.

(1976), não apresentam a justificação como um componente necessário do conhecimento : o conhecimento é uma crença com uma determinada origem causal, que não é ela mesma uma justificação para a crença.<sup>24</sup> Uma proposta mais claramente oposta à análise tripartite consiste em que considerar o conhecimento não é analisável em componentes independentes, mas é um conceito primitivo.<sup>25</sup> Na proposta de Tim Williamson, a fiabilidade é um traço do conhecimento, mas o conhecimento ele mesmo não pode ser decomposto em elementos mais primitivos, como o faz a definição (D).<sup>26</sup> O contraste entre a crença meramente verdadeira e o conhecimento traz à tona a robustez deste último, em oposição à fragilidade da primeira.<sup>27</sup> Ana vai fazer uma excursão na floresta e vê um fruto azul. Ela sabe que se trata de um fruto venenoso e procura outra coisa para se alimentar. Arnaldo tem a mesma crença verdadeira, porque pensa que todo fruto azul é venenoso. Arnaldo vê então um pé de araquá azul e se dá conta que nem todo fruto azul é tóxico. Arnaldo muda de opinião e passa a ter a crença falsa que o primeiro fruto não é venenoso. Ana, por outro lado, será sensível ao que de fato mostra se o fruto em questão é venenoso e não muda sua crença. O conhecimento de Ana é mais estável que a crença verdadeira de Arnaldo.<sup>28</sup>

A robustez do conhecimento, i.e., o fato de ele ser mais estável através de mudanças de contextos do que a crença meramente ver-

<sup>24</sup> Sobre o desenvolvimento inicial de Goldman, que o leva à postular o fiabilismo como teoria da justificação, ver Kornblith 2008: 6-8. Kornblith ele mesmo é um exemplo de um teoria do conhecimento sem justificação – ver Kornblith 2002; 2008.

<sup>25</sup> Williamson 2000; ver também Dretske 1981, chap. 4. Para um debate sobre este tema, ver Williamson 2014 e Dougherty e Rysiew 2014. Sobre a relação entre a proposta de Williamson e o projeto de análise conceitual, ver Cassam 2009.

<sup>26</sup> Williamson 2000: 41.

<sup>27</sup> “O conhecimento é superior à mera crença verdadeira porque, sendo mais robusto diante das novas evidências, facilita mais a ação distante no tempo. O resto permanecendo igual, dada a sensibilidade racional a novas evidências, o conhecimento presente torna uma crença futura verdadeira mais provável do que uma mera crença verdadeira presente.” Williamson 2000: 101.

<sup>28</sup> Ver exemplos semelhantes em Williamson 2000: 87, 101.

dadeira, é um tipo de fiabilidade. Ana rastreia a comestibilidade do fruto em diferentes situações, ao passo que Arnaldo mais facilmente passará de uma crença verdadeira a uma crença falsa. Na medida em que consideramos um conjunto maior de situações, os comportamentos dos dois divergem e a ação de Ana será mais adequada num número maior de casos.

O papel do conhecimento na explicação da ação explora um tipo de fiabilidade. Se num tempo  $t$ , a partir de uma base  $b$  uma pessoa sabe que  $p$ , e num tempo  $t^*$  próximo o bastante de  $t$ , a partir de uma base  $b^*$ , próxima o bastante de  $b$ , esta pessoa acredita que  $p^*$ , próxima o bastante de  $p$ , então  $p^*$  deve ser verdadeira.<sup>29</sup>

Reencontramos os traços centrais do fiabilismo: o rastreamento da verdade em diferentes situações, ou, mais precisamente, em situações similares à situação de avaliação da crença, nas quais alternativas relevantes são excluídas. Esta capacidade não é infalível, mas funciona em situações similares à situação de avaliação da crença. Não voltaremos ao debate acerca do caráter primitivo do conhecimento, suscitado sobretudo pela proposta de Williamson. A diferença o fiabilismo como teoria da justificação e o fiabilismo como teoria do conhecimento repousa, em boa medida, sobre o uso do termo ‘justificação’ – voltaremos a este tema abaixo.

### 3 As críticas ao Fiabilismo

Há três críticas centrais ao fiabilismo. As duas primeiras procuram mostrar que a normatividade epistêmica exige que se leve em conta a perspectiva do sujeito. A terceira crítica explora a dificuldade da determinação do tipo sob o qual caem processos particulares de obtenção de crenças que são relevantes para sua avaliação.

A primeira crítica é conhecida como o novo problema do gênio maligno.<sup>30</sup> Suponhamos um mundo  $m'$  no qual  $S$  tenha as mesmas crenças que no mundo efetivo, com a mesma confiança, intensidade

<sup>29</sup> Williamson 2000: 102. No exemplo do fruto azul, a base da crença é a experiência visual.

<sup>30</sup> Este argumento foi proposto inicialmente por Lehrer e Cohen 1983 e Cohen 1984. Sobre o novo problema do gênio maligno, ver Littlejohn.

e clareza, a partir de experiências similares. No entanto,  $m'$  é controlado por um gênio maligno, capaz de produzir em  $S$  experiências que quase sempre levam a crenças falsas. As crenças de  $S$  não são produzidas por um processo fiável. Elas parecem contudo ser justificadas:  $S$  é epistemicamente responsável, por tudo o que ele sabe, suas crenças foram produzidas por um processo fiável. A justificação é uma noção normativa e como, na situação descrita,  $S$  não age de maneira epistemicamente irresponsável, ele está justificado em suas crenças.<sup>31</sup> O novo problema do gênio maligno não visa uma conclusão cética, como o argumento cartesiano, mas busca mostrar como justificação e verdade são independentes; é um argumento precisamente anti-fiabilista. O fiabilismo e o internismo têm vereditos opostos acerca da justificação do sujeito em  $m'$ . O termo 'justificação' não parece ter um significado teoricamente neutro e claro o suficiente para estabelecer que apenas um destes usos é correto. O que está em jogo aqui não é o uso mais adequado do termo, mas o que pode desempenhar o papel da terceira condição no conhecimento. O ponto central desta crítica é que o elemento normativo é constitutivo do conhecimento e só pode ser corretamente aplicado com critérios internistas, i.e., segundo o que é apreendido pelo sujeito ou, numa outra versão, por traços supervenientes sobre propriedades mentais do sujeito.

A primeira crítica busca mostrar que a fiabilidade não é uma condição necessária para a justificação. A fiabilidade do processo pode não ser tampouco uma condição suficiente, como é mostrado no problema da meta-incoerência.<sup>32</sup> Laurence Bonjour propõe uma série de casos nos quais videntes acreditam corretamente que o presidente dos Estados Unidos está em Nova York. Numa das situações propostas, Norman é um vidente que não tem razões a favor ou contra sua clarividência. As crenças que obtém deste modo são, no entanto, em geral verdadeiras. Um dia, sem razões para tal, ele acredita que o presidente dos Estados Unidos está em Nova York. Segundo o fiabilismo, ele está justificado em ter tal crença, mesmo se não tem razão alguma para acreditar naquilo que seu poder de clarevidência o leva a acreditar. Como Norman poderia acreditar de maneira justificada

<sup>31</sup> Lehrer e Cohen 1983: 193; Cohen 1984: 282.

<sup>32</sup> Sigo aqui o nome proposto por Sosa 1991: 132.



na sua crença, se não tem razões para julgar seus poderes fiáveis?<sup>33</sup> Se a hipótese de mundos demoníacos mostra que a fiabilidade do processo não é necessária à justificação da crença, esta crítica mostra que ela não é suficiente: um processo fiável do qual o sujeito não tem consciência não pode ser, para este sujeito, o que justifica sua crença.<sup>34</sup>

Uma mostra da dificuldade encontrada pelo fiabilismo está no modo como Goldman é levado, ao longo de seus textos, a incluir um elemento interno. Em “What is a justified belief”, ele diz que para explicar o que entendemos normalmente por justificação, devemos nos reportar às nossas crenças acerca da fiabilidade dos processos.<sup>35</sup> Isto equivale a considerar a fiabilidade como um elemento daquilo que, do ponto de vista do sujeito, é levado em conta na avaliação de uma crença, isto é, um elemento interno. Nossas crenças acerca do caráter fiável de nossos procedimentos de obtenção de crenças podem se mostrar falsas, o que nos levaria a uma imagem internista da justificação – este é o caso, por exemplo, do mundo demoníaco descrito no novo argumento do gênio maligno. Um movimento análogo aparece em *Epistemology and Cognition* (1986). Para evitar mundos demoníacos, Goldman acrescenta a exigência que julgamos justificadas as crenças que foram produzidas por processos fiáveis em mundos normais. Mas como definir mundos normais? Um mundo normal é aquele consistente com nossas crenças gerais sobre o nosso mundo.<sup>36</sup> Mais uma vez, a referência a nossas crenças torna o processo de jus-

<sup>33</sup> Bonjour 1980.

<sup>34</sup> “Como pode o fato de uma crença ser produzida de maneira fiável (ou qualquer outro tipo de fato que faz com que uma crença tenha muitas chances de ser verdadeira) tornar minha aceitação desta crença racional e responsável, quando este fato é ele mesmo inteiramente inacessível para mim?” Bonjour, In Bonjour e Sosa 2003: 27.

<sup>35</sup> “O que realmente queremos é uma *explicação* da razão pela qual consideramos, ou considerariamos, que certas crenças estão justificadas e outras não. Tal explicação tem de se referir às nossas *crenças* sobre a fiabilidade, e não a  *fatos* efetivos. A razão pela qual *consideramos* que as crenças estão justificadas é que foram formadas pelo que *cremos* serem processos fiáveis de formação de crenças. As nossas crenças sobre que processos de formação de crenças são fiáveis podem ser errôneas, mas isso não afeta a adequação da explicação.” Goldman 2011 (1992: 121).

<sup>36</sup> Goldman 1986: 107

tificação dependente de um elemento, que pode ser falsificado num mundo demoníaco, ou simplesmente pouco cooperativo.<sup>37</sup>

Uma outra importante crítica ao fiabilismo é o chamado problema da generalidade: como cada ocorrência particular de um processo instancia um número indeterminado de tipos de processos, a fiabilidade de qual tipo de processo é relevante para determinar se esta ocorrência particular é um caso de conhecimento ou não?<sup>38</sup> Ana olha para a sala de sua casa e vê um gato preto. A crença que resulta desta experiência é um exemplo de uma crença perceptual obtida em condições favoráveis: ela instancia um processo que tende a produzir mais crenças verdadeiras do que falsas. No entanto, é também uma crença obtida por um processo que leva precisamente à crença verdadeira que há um gato em sua sala naquela quarta-feira à tarde, que é uma delimitação muito restrita do tipo relevante para a avaliação. Ele ainda uma instância do processo que leva a crenças acerca de gatos a partir de uma experiência visual que faz pensar em um gato, exemplificado também pela crença obtida por Arnaldo a partir da borra de café, o que é uma tipificação muito ampla do processo em questão, que perde diferenças epistemicamente relevantes. Trata-se ainda de uma crença adquirida numa terça-feira chuvosa por uma pessoa sentada num sofá. Cada processo efetivo de obtenção de crenças instancia um número indeterminado de tipos processos. Qual destes tipos é relevante para a determinação da fiabilidade de sua crença? Não há uma maneira não arbitrária de selecionar o tipo exemplificado por um dado processo de obtenção de crenças e, sem esta seleção, notam Earl Conee e Richard Feldman, o fiabilismo é “radicalmente incompleto”.<sup>39</sup>

A resposta fiabilista mais influente ao problema da generalidade se deve a William Alston.<sup>40</sup> O argumento que o estabelece, diz Alston, depende da assunção que não há fatos psicológicos objetivos

<sup>37</sup> Uma excelente análise do percurso de Goldman encontra-se no capítulo 7 de Haack 1993.

<sup>38</sup> O problema da generalidade foi apontado por Goldman já no seu artigo de 1979; Goldman 1992: 115-116.

<sup>39</sup> Conee e Feldman 1998.

<sup>40</sup> Alston 1995.

“que selecionem um único tipo como sendo aquele do qual um dado processo particular é uma ocorrência.”<sup>41</sup> A multiplicação de tipos sob os quais cai um determinado processo vale para qualquer tipo de classificação: Arnaldo pertence aos conjuntos de objetos que pesam mais de 70 quilos, de objetos que de longe se parecem com moscas e de coisas mencionadas neste texto. O que explica no entanto o que ele é fundamentalmente, diz Alston, é o fato de ele ser um ser humano, o tipo natural ao qual ele pertence. Processos psicológicos são como tipos naturais, há portanto uma questão de fato sobre qual tipo um dado processo particular de obtenção de crenças instancia.<sup>42</sup> A obtenção de crenças “envolve a ativação de uma determinada função psicologicamente realizada.”<sup>43</sup> Em cada processo particular de formação de crenças, “há apenas um “tipo”, i.e., apenas uma função realizada, que é *causalmente operativa* (...), uma vez excluídos os casos de sobredeterminação.”<sup>44</sup> Esta resposta repousa sobre um traço central do fiabilismo, a saber, a fundamentação da justificação na explicação da origem causal da crença. Como veremos, a proposta de Alston pode ser expandida num quadro de naturalização da epistemologia. As respostas ao novo argumento do gênio maligno e ao problema da meta-incoerência irão indicar diferentes caminhos de desenvolvimento do fiabilismo, a via da epistemologia da virtude e o caminho da naturalização da epistemologia.

#### 4 Uma resposta fiabilista: a epistemologia da virtude

Uma motivação central para as duas primeiras críticas internistas ao fiabilismo é o abandono do ponto de vista do sujeito e a consequente eliminação da normatividade da epistemologia. Chamemos ‘fiabilismo genérico’<sup>45</sup> a tese que o conhecimento é uma crença verdadeira resultante de um processo que leva a mais crenças verdadeiras do que a crenças falsas. Algo deve ser acrescentado ao fiabilismo genérico

<sup>41</sup> Alston 1995: 5.

<sup>42</sup> Alston 1995: 10-11.

<sup>43</sup> Alston 1995: 12.

<sup>44</sup> Alston 1995: 23.

<sup>45</sup> Sosa 1991: 131.

para acomodar a perspectiva do sujeito; a epistemologia da virtude, pelo menos em algumas de suas versões, é a via mais usual de resposta a este problema no quadro fiabilista.<sup>46</sup>

A formulação inicial da epistemologia da virtude se deve a Ernest Sosa.<sup>47</sup> Sua tese central é que uma crença verdadeira é conhecimento se ela foi causada por uma disposição estável do sujeito que tende a produzir crenças verdadeiras sobre um determinado tipo de proposição em determinados ambientes, i.e., por uma virtude intelectual.<sup>48</sup> Uma crença cuja verdade é assim explicada é uma crença apta:

Podemos distinguir entre a acurácia de uma crença, isto é, sua verdade; sua habilidade, isto é, sua manifestação de virtude epistêmica ou competência; e sua aptidão, isto é o fato de ser verdadeira *porque* competente.<sup>49</sup>

A acomodação da perspectiva do sujeito demanda um passo suplementar. Sosa distingue o conhecimento animal e conhecimento reflexivo.<sup>50</sup> O conhecimento animal é simplesmente uma crença apta, i.e., uma crença cuja verdade é explicada pelo exercício das habilidades cognitivas do sujeito. O conhecimento reflexivo é uma crença endossada pelo sujeito de maneira apta, i.e., uma crença acerca de outra crença que resulta do exercício de sua competência na

<sup>46</sup> A epistemologia da virtude é em geral dividida em dois campos, teorias fiabilistas, como as de Sosa e Greco, e teorias responsabilistas, como a defendida por Zagzebski. Para uma comparação entre estes dois tipos de epistemologia da virtude, ver Greco e Turri 2015. Neste texto, vou me concentrar na primeira família de teorias.

<sup>47</sup> “The Raft and the Pyramid” (1980) em Sosa 1991: 189-190.

<sup>48</sup> Sosa 1991: 286.

<sup>49</sup> Sosa 2013a : 32 (2007 : 23) ; tradução ligeiramente modificada.

<sup>50</sup> A primeira versão desta distinção é a seguinte: “Uma pessoa tem um *conhecimento animal* sobre o seu ambiente, seu passado e sua própria experiência se seus julgamentos e crenças sobre este temas são respostas diretas ao impacto de seu ambiente, seu passado e sua experiência – e.g., através da percepção ou da memória –, com pouco ou nenhum benefício da reflexão ou da compreensão. Uma pessoa tem um *conhecimento reflexivo* se seu julgamento ou crença manifesta não apenas uma resposta direta ao fato conhecido, mas também a compreensão do seu lugar num todo mais amplo que inclui sua própria crença e o conhecimento de como ela foi produzida.” Sosa 1991: 240.

avaliação de seus próprios estados epistêmicos.<sup>51</sup> O conhecimento reflexivo demanda que tenhamos “uma visão de nós mesmos – de nossas crenças, de nossas faculdades e da nossa situação – à luz da qual possamos ver as fontes de nossas crenças como suficientemente fiáveis”:<sup>52</sup>

(...) podemos, de maneira geral, distinguir por um lado o conhecimento *animal*, que requer apenas o rastreamento da natureza, e por outro o conhecimento *reflexivo*, que requer também a consciência de *como se sabe*, de uma maneira que exclui a não confiabilidade de nossas faculdades.<sup>53</sup>

Estes diferentes exercícios de virtudes cognitivas resultam em diferentes tipos de justificação:

A justificação *não reflexiva*, justificação animal, não depende de nenhum endosso reflexivo; de nenhum endosso da *fiabilidade específica* da nossa base (...) *Nem* depende essa justificação racional de nenhum endosso da *fiabilidade genérica* da nossa base. (...) A justificação racional *reflexiva*, em contraste, é adquirida pelo menos em parte por meio de endosso racional: *ou* por endosso da *fiabilidade específica* de nossa base (...), *ou* pelo endosso da *fiabilidade genérica* de nossa base.<sup>54</sup>

É possível se explicar o crédito que o sujeito tem pela verdade de suas crenças sem postular uma capacidade metateórica, como parece exigir Sosa. Segundo John Greco, o crédito do sujeito pelo conhecimento reside no exercício cuidadoso de suas habilidades intelectuais, que explica porque sua crença é verdadeira. Uma crença cuja verdade é explicada deste modo é subjetivamente justificada.<sup>55</sup> Contrariamente à teoria de Sosa, a abordagem de Greco não demanda o passo suplementar do conhecimento reflexivo, ou pelo menos nem sempre é necessário que o sujeito tenha crenças sobre as próprias crenças

<sup>51</sup> “O conhecimento animal é uma crença apta de primeira ordem. O conhecimento reflexivo é uma crença animal endossada de maneira apta pelo sujeito” Sosa 2011: 12.

<sup>52</sup> Sosa 2013b: 139 (2009: 147).

<sup>53</sup> Sosa 2013b: 182 (2009: 199).

<sup>54</sup> Sosa 2013b: 215-216 (2009: 238-239).

<sup>55</sup> “Uma crença *p* é subjetivamente justificada para *S* se, e apenas se, *p* se baseia em disposições cognitivas que *S* manifesta quando pensa cuidadosamente.” Greco 2000: 190.

para estar justificado. A fiabilidade de nossas crenças não exige, por parte do sujeito, crenças acerca dos próprios processos de obtenção de crenças, não mais do que um atleta precisa ter crenças acerca das próprias habilidades para ter crédito por seus feitos – nos termos Greco, podemos preservar a virtude epistêmica sem ter uma perspectiva sobre nossos próprios estados.<sup>56</sup> Entretanto, como veremos na resposta ao problema da meta-incoerência, Greco é levado a uma posição próxima da de Sosa.

A epistemologia da virtude permite abordagens mais satisfatórias do novo problema do gênio maligno e do problema da meta-incoerência do que o fiabilismo genérico. A resposta de Sosa se faz pela distinção entre o mundo de exercício e o mundo de avaliação de crenças:

J-A (para todo  $m$ ) (uma crença  $B$  está justificada-A em  $m$  apenas se  $B$  foi adquirida em  $m$  pelo exercício de uma ou mais virtudes intelectuais que são virtuosas em  $m$ )

J-H (para todo  $m$ ) (uma crença  $B$  está justificada-H em  $m$  apenas se  $B$  foi adquirida em  $m$  pelo exercício de uma ou mais virtudes intelectuais que são virtuosas no nosso mundo efetivo).<sup>57</sup>

Virtudes são definidas correlativamente como virtudes-A e virtudes-H. Na primeira definição, crenças são justificadas relativamente a virtudes que produzem um alto índice de crenças verdadeiras no mundo no qual são adquiridas; na segunda definição, virtudes que justificam crenças produzem mais crenças verdadeiras no nosso mundo efetivo, mesmo se não forem virtuosas, i.e. não produzirem um alto índice de crenças verdadeiras, no mundo no qual elas são exercidas.

Na situação imaginada pelo novo problema do gênio maligno, o sujeito está justificado-H, mas não justificado-A:

À vítima do gênio maligno cartesiano *não* parece faltar a justificação *ordinária*, pois suas crenças ainda derivam de fontes que nós reconhecemos como capazes de conferir justificação, nomeadamente, a experiência sensorial, a memória etc. O ambiente muda radicalmente, mas a

<sup>56</sup> Greco 2000: 141. Para um contraste entre as posições de Greco e Sosa sobre este ponto, ver Greco 2004 e a resposta de Sosa 2004: 290-292.

<sup>57</sup> Sosa 2003: 156-157.

vítima mantém e utiliza um repertório de virtudes intelectuais.<sup>58</sup>

Esta distinção mostra que a justificação é sempre relativa ao exercício virtuoso de uma capacidade cognitiva em algum ambiente. O novo argumento do gênio maligno não rompe esta conexão, mas explora a separação possível entre o ambiente de exercício e o ambiente de avaliação de uma dada crença: há um sentido no qual o sujeito, nesta circunstância, está justificado porque suas crenças resultam do exercício de capacidades cognitivas que são virtuosas no ambiente a partir do qual é feita a avaliação. A distinção cobre a justificação não reflexiva, pois ela é independente da perspectiva do sujeito sobre si mesmo. Ela explica no entanto porque o avaliador, de um ponto de vista externo, pode julgar que, num mundo demoníaco, o sujeito ainda está justificado.

O desafio da meta-incoerência recebe uma resposta relativamente direta se aceitarmos que há de fato conhecimento apenas quando temos conhecimento reflexivo. O que falta ao vidente Normam na história proposta por BonJour é precisamente a perspectiva sobre seus próprios atos, ele não tem portanto propriamente conhecimento. Se pensarmos contudo que a virtude epistêmica não demanda a perspectiva sobre seus próprios atos, i.e., não demanda crenças sobre as próprias crenças, como responder a esta crítica? Na resposta de Greco, se seu poder como vidente estiver em oposição com outras crenças suas, a mera fiabilidade não constitui uma agência epistêmica virtuosa. Se, no entanto, a fiabilidade de sua clarividência for compatível com suas crenças acerca de seus poderes cognitivos – o que é o caso se não houver oposição entre ambos os conjuntos de crenças, então ele é epistemicamente virtuoso. As crenças verdadeiras que resultarem do exercício de sua clarividência serão neste caso conhecimento.<sup>59</sup> O argumento da meta-incoerência exige a integração de nossas crenças e, de certo modo, a diferença entre Sosa e Greco é sobretudo uma questão de grau.<sup>60</sup> Esta última objeção mostra que

<sup>58</sup> Sosa 2003: 156.

<sup>59</sup> Greco 2000: 141-142.

<sup>60</sup> Ver integração implícita das crenças, e.g., Sosa 2009: 139. Sosa propõe uma extensão da noção de reflexão para incluir processos sub-pessoais em Sosa 2015: 187-188, em resposta a Carvalho e Williges 2015: 157-158.

a epistemologia deve se aproximar do internismo, para explicar a racionalidade das crenças do sujeito.<sup>61</sup>

## 5 Uma outra resposta fiabilista: o caminho para a naturalização

Uma outra resposta fiabilista não busca a acomodação da perspectiva do sujeito, mas, ao contrário, nega que ela seja necessária. O primeiro passo nesta direção consiste, mais uma vez, na distinção entre dois tipos de justificação. Kent Bach distingue a justificação do sujeito e a justificação de uma crença: enquanto a justificação do sujeito “reside na qualidade de sua ação epistêmica”, que o leva à crença de maneira racional e responsável, uma crença “pode estar justificada mesmo sem qualquer ação por parte do sujeito, como no caso de crenças formadas automaticamente ou por processos rotineiros, sem qualquer deliberação”.<sup>62</sup> Mais uma vez, não se trata de uma mera questão terminológica, que poderia ser resolvida pela adoção de um ou outro sentido de ‘justificação’. O problema é o de saber se a separação entre conhecimento e crenças meramente verdadeiras requer alguma ação do sujeito, em particular a ação de apresentar uma justificação para sua crença. Fred Dretske expõe a diferença entre internistas e externistas que se apresenta aqui do seguinte modo:

O debate entre externistas e internistas em epistemologia pode ser visto como um desacordo sobre a existência de direitos epistêmicos (de se acreditar em algo) sem os deveres ou obrigações correspondentes (de se justificar aquilo em que se acredita).<sup>63</sup>

A tese externista é que o sujeito pode ter direitos epistêmicos sem deveres correspondentes e, em particular, sem o dever de se justificar em ter uma crença determinada. A teoria fiabilista assim desenvolvida será mais aberta ao conhecimento de animais incapazes de entrar no jogo de dar e receber razões, como animais não linguísticos e humanos pré-linguísticos.<sup>64</sup> Muitas de nossas crenças resultam de

<sup>61</sup> Sosa 2009: 44.

<sup>62</sup> Bach 1985: 251. Ver também Kornblith 2002; 2008.

<sup>63</sup> Dretske 2000: 591.

<sup>64</sup> Este ponto é destacado, por exemplo, por Dretske 1991.



raciocínios automáticos, segundo a expressão de Bach:

... uma teoria razoável sobre crenças justificadas (e assim sobre o conhecimento) deve levar em conta limitações reais de nossos raciocínios cotidianos. Filósofos tendem a focar sobre os raciocínios mais explícitos e deliberados. No entanto, tais raciocínios são excepcionais: a maior parte dos raciocínios que nos dão conhecimento é amplamente não explícita. Eles são o que, na inteligência artificial, é chamado *raciocínio automático* [*default reasoning*], cada passo se baseia em alguma generalização ou estereótipo, que é abandonado apenas se ocorrer o pensamento de uma alternativa ou uma razão contrária. (...) Este fenômeno pervasivo pode levar a crenças justificadas, na medida em que ele é fiável. Sua fiabilidade depende (...) de quão confiáveis nós somos em saber quando pensar duas vezes.<sup>65</sup>

Mas o que desencadeia uma razão para bloquear um mecanismo em geral fiável? A pista vem, de maneira talvez surpreendente, de uma explicação de Sosa:

Um ser dotado de razão monitora automaticamente sua informação de fundo e suas entradas sensoriais indicando evidências contrárias e escolhe automaticamente a hipótese mais coerente, mesmo quando responde da maneira mais direta aos estímulos sensoriais. Mesmo quando a resposta aos estímulos é a mais direta, o sujeito mudaria sua resposta se escutasse ou visse sinais de testemunhos contrários críveis. As crenças de uma animal *racional* parecem assim nunca resultar da introspecção, da memória ou da percepção *isoladas*, pois a razão é sempre pelo menos um parceiro silencioso que vigia outros dados relevantes, um parceiro silencioso cujo *silêncio* mesmo é uma causa que contribui para a formação da crença.<sup>66</sup>

O monitoramento automático não envolve a reflexão sobre o que se conhece, mas a disposição do sujeito de reconhecer situações desfavoráveis de exercício de uma dada capacidade cognitiva.<sup>67</sup> Se esta des-

<sup>65</sup> Bach 1985: 254.

<sup>66</sup> Sosa 1991: 240.

<sup>67</sup> John Turri sugere que uma maneira de se desenvolver a epistemologia de dois níveis de Sosa é a de recorrer à distinção entre dois sistemas de pensamento; Turri 2013: 158-159. Segundo a teoria de Daniel Kahneman e Amos Tversky à qual Turri se refere, o sistema 1 dá respostas automáticas, não conceituais, sobre as quais se baseiam as avaliações meta-representacionais do sistema 2; para uma apresentação do sistema dual de pensamento, ver Kahneman 2012 (2011). Uma versão mais complexa de uma teoria da mente em diferentes níveis se encontra Stanovich 2011; em relação aos temas que nos concernem, ver em particular pp. 61-80. Comentando esta passagem de Sosa, Kornblith diz que a “sensibilidade a

criação do conhecimento reflexivo não requer a representação, pelo sujeito, das condições nas quais ele deve suspender o funcionamento habitual dos mecanismos cognitivos, ela demanda mecanismos cognitivos sub-pessoais que respondam ao funcionamento dos mecanismos cognitivos mesmos do sujeito, i.e., mecanismos metacognitivos.

A postulação de processos de monitoramento independentes de capacidades metarepresentacionais é compatível com teorias externistas do conhecimento. Os mecanismos que tornam as crenças justificadas podem não ser apreendidos pelo sujeito, como é o caso dos mecanismos sub-pessoais postulados aqui. Como observa Greco sobre Sosa, capacidades metarepresentacionais exigem que sujeito seja capaz de se representar os processos de produção de crenças, o que parece ser uma demanda muito forte. Uma outra explicação destes processos consiste na postulação de uma “metacognição procedimental”: a comparação entre expectativas de um agente, ou de seus sistemas perceptuais, e as pistas encontradas no desenrolar de sua experiência não se faz, ou pode não se fazer, no nível pessoal. Há mais de um modelo proposto para explicar a avaliação da incerteza a partir das características informacionais da atividade cerebral que são parte da explicação do que desencadeia o processo de avaliação das próprias crenças por parte do sujeito.<sup>68</sup>

A naturalização da epistemologia se estende aos diferentes aspectos dos sistemas cognitivos, não apenas à metacognição. Assim, em muitos casos, a relativização da fiabilidade de processos cognitivos a contextos é explicada por diferentes mecanismos elucidados pelas ciências cognitivas. Nosso sistema visual, por exemplo, inclui limitações às interpretações a que pode chegar a partir das entradas sensoriais que recebe que refletem traços gerais do mundo em que vivemos.<sup>69</sup> O que se esboça aqui é uma ponte da epistemologia para

evidências contrárias” não exige metarepresentação; Kornblith 2004: 128.

<sup>68</sup> Proust 2012. Para a sugestão de um quadro naturalista geral que acomoda o monitoramento dos próprios estados epistêmicos, ver Millikan 1984.

<sup>69</sup> Como explica Zenon Pylyshyn, a partir do trabalho de David Marr, pode-se mostrar como, para a visão, “tudo o que é necessário é que as computações realizadas pelo processamento primário incorpore (sem representar de maneira explícita ou realizar inferências) algumas condições muito gerais para as interpretações que pode fazer. Estas condições não garantem a interpretação correta de todos

as ciências cognitivas, que se segue do foco no processo de produção de crenças. Na medida em que os processos de aquisição e manutenção de crenças são relevantes para a epistemologia, devemos nos preocupar com a plausibilidade das teses psicológicas assumidas. Como argumentam Henderson e Horgan,<sup>70</sup> o projeto epistemológico tradicional acomoda esta exigência ao postular o isomorfismo entre estados sub-pessoais epistemologicamente relevantes e estruturas justificacionais definidas pelo conteúdo proposicional e portanto articuláveis, mesmo quando não efetivamente articuladas, pelo sujeito. No entanto, este suposto isomorfismo não é empiricamente plausível. Os diferentes modelos de implementação dos mecanismos de comparação entre valores perceptuais esperados e valores encontrados na experiência, que fornecem uma explicação para o que desencadeia o bloqueio e a reavaliação de raciocínios automáticos, são exemplos de processos epistemologicamente relevantes que não são isomórficos às razões que um sujeito pode apresentar para suas crenças metateóricas e que não são portanto articuláveis pelo sujeito.<sup>71</sup>

A pressão das duas primeiras críticas ao fiabilismo diminui muito num quadro naturalista. Em relação ao novo argumento do gênio maligno, não é preciso postular que explicações que valem no mundo efetivo e que resultam da história evolutiva da nossa espécie se estendam para mundos diferentes do nosso. O argumento da meta-incoerência também não parece ter muito força, uma vez que se abandona a necessidade da reflexão para o conhecimento. Pode-se acrescentar que o monitoramento das próprias crenças, o que é um parte do que se demanda da integração cognitiva, pode ser obtido sem meta-representação, o caso de Norman deve então ser redescrito para definir se, nesta história, ele tem a capacidade de monitoramento. No

os estímulos. (...) Tudo o que é necessário é que as condições naturais levem à interpretação correta sob condições específicas que ocorrem frequentemente no nosso tipo de mundo físico.” Pylyshyn 2003: 96. Estas condições não dependem de uma teoria representacional da visão, ou, mais geralmente, da percepção; ver, por exemplo, a descrição de pressuposições (não doxásticas) relativas ao mundo incorporadas pelas diferentes modalidades sensoriais numa teoria enativa da percepção em Bridgeman 2002.

<sup>70</sup> Henderson e Horgan 2000.

<sup>71</sup> Sobre estes modelos, ver Proust 2012 e 2013: 13-78.

entanto, talvez a consequência mais importante da naturalização seja o afastamento de teorias construídas a partir de nossos julgamentos ordinários acerca do conhecimento.<sup>72</sup> Tanto o novo argumento do gênio maligno quando o desafio da meta-incoerência pedem nossos julgamentos intuitivos sobre situações contrafactuais bem distantes do mundo efetivo, e é precisamente o peso deste julgamento que diminui aqui.

A resposta de Alston ao problema da generalidade pode ser estendida num quadro naturalista. Uma resposta consiste em dizer que o conhecimento é um tipo natural e que portanto o mundo mesmo define o tipo que cada processo particular de obtenção de crenças exemplifica.<sup>73</sup> Não é preciso no entanto postular que haja um único tipo relevante instanciado por cada processo cognitivo. Como observam Henderson e Horgan, para um dado fenômeno cognitivo, há mais de uma explicação possível:

(...) dada a questão exata que se está a responder, pode-se explicar o veredito perceptual em termos de certos objetos salientes no ambiente do agente, ou em termos da formação de fundo do agente e das sensibilidades resultantes, ou em termos de *primings* contextuais, e assim por diante.<sup>74</sup>

Uma consequência direta da pluralidade de explicações é a pluralidade resultante de tipos sob os quais cai um processo cognitivo. Não se trata de uma pluralidade arbitrária, mas está ligada ao tipo de pergunta que se faz acerca de um dado processo. Alguns dos exemplos propostos por Alston apontam na direção de explicações científicas

<sup>72</sup> Em 1976, Goldman ainda pretende explicar as aplicações do termo 'saber'; Goldman 1992: 102. Dretske pensa antes numa "coordenação" entre o conceito ordinário de conhecimento e noções "técnicas", como o conceito de informação, ou ainda uma "ponte" entre a análise conceitual e as ciências cognitivas (Dretske 1979: 91-92). Neste quadro teórico, o termo 'conhecer' e cognatos, não mais do que 'justificação' e expressões derivadas, não são estáveis o bastante para oferecer a base da epistemologia, o que permite a acomodação da posição de Dretske em 1979, separando conhecimento, entendido numa chave causal, e fiabilidade (Dretske 1979: 86-92).

<sup>73</sup> Kornblith 2002: 63-69.

<sup>74</sup> Henderson e Horgan 2011: 90.

de nossos processo cognitivos.<sup>75</sup> O problema é resolvido porque a seleção do tipo relevante seguirá a explicação aceita para um dado processo, e explicações não se multiplicam de maneira arbitrária.

## 6 Conclusão

Para teorias do conhecimento externistas, a distinção entre o conhecimento e crenças meramente verdadeiras não depende daquilo a que o sujeito a quem se atribui conhecimento tem acesso consciente. Teorias fiabilistas são um tipo de resposta externista a esta questão: uma crença verdadeira é conhecimento se ela resulta de um processo que, num dado tipo de contexto e para um determinado tipo de proposição, tende a produzir mais crenças verdadeiras do que crenças falsas. Neste quadro, o que torna uma crença verdadeira conhecimento pode escapar ao sujeito por duas razões. Inicialmente, não temos acesso introspectivo aos processos sub-pessoais de produção de nossas crenças. Além disto, que determinados processos tendam a levar a crenças verdadeiras num determinado tipo de ambiente também envolve um elemento externo ao sujeito, que é precisamente o fato de as crenças tenderem a ser verdadeiras, e não falsas.

Ao focar no processo de produção de crenças, o fiabilismo irá se interessar pela compreensão sobre o modo como crenças se originam no sujeito. Esta compreensão irá aproximar o fiabilismo das ciências cognitivas. Este é o caso em projetos naturalistas, mas também da razão como “parceira silenciosa” em Sosa, que demanda uma explicação em termos de mecanismos sub-pessoais. Se este caminho mantém a fiabilidade de processos cognitivos como a marca distintiva do conhecimento, ele se afasta do projeto tradicional da epistemologia em torno do qual o fiabilismo se construiu e, em particular, sua dimensão normativa. Uma outra resposta é associar a fiabilidade a virtudes epistêmicas do sujeito, que podem ser o objeto de avaliação epistêmica. Se a epistemologia das virtudes parece ser o melhor caminho para se manter o fiabilismo no quadro tradicional da epistemologia, ela exhibe ainda a tensão entre a perspectiva metateórica que parece ser uma demanda da justificação do sujeito e teorias realistas

<sup>75</sup> Ver Alston 1995: 13-15.

acerca dos mecanismos de produção de crenças.<sup>76</sup>

Ernesto Perini Santos  
Departamento de Filosofia  
UFMG

### Referências

- Alston, William P. How to Think about Reliability. *Philosophical Topics* **23**: 1-29.
- Bach, Kent. 1985. A Rationale for Reliabilism. *The Monist* **68**: 246-263.
- Bonjour, Laurence. 1980. Externalist Theories of Empirical Knowledge. In French, P. A., T. E. Uehling e H. K. Wettstein. *Midwest Studies in Philosophy: Studies in Epistemology*. University of Minnesota Press.
- Brandom, Robert. 2000. *Articulating Reasons*. Cambridge, Mass.: Harvard University Press.
- Bridgeman, Bruce. 2002. The Grand Illusion and Petit Illusions. *Journal of Consciousness Studies* **9**: 29-34.
- Carvalho, Eros Moreira e Flavio Williges. 2015. Sosa on Animal Knowledge and Emotions. *Analytica* **19**: 145-160.
- Cassam, Quassim. 2009. Can the Concept of Knowledge be Analysed? In Greenough, P. e D. Pritchard (eds.) *Williamson on Knowledge*. Oxford University Press.
- Chisholm, Roderick. 1989. *Theory of Knowledge*. 3rd Edition. Prentice-Hall.
- Cohen, S. 1984. Justification and Truth. *Philosophical Studies* **46**: 279-96.
- Conee, Earl e Richard Feldman. 2004. *Evidentialism – Essays in Epistemology*. Oxford University Press.
- Craig, Edward. 1990. *Knowledge and the State of Nature*. Oxford University Press.
- Dretske, Fred. 1970. Epistemic Operators. *The Journal of Philosophy*. **67**: 1007-1023.
- Dretske, Fred. 1981. *Knowledge and the Flow of Information*. MIT Press.
- Dretske, Fred. 1988. *Explaining Behavior – Reasons in a World of Causes*. MIT Press.
- Dretske, Fred. 1991. Two Conceptions of Knowledge: Rational vs. Reliable Belief. *Grazer Philosophische Studien* **4**: 15-30.
- Dretske, Fred. 2000. Epistemic Rights without Epistemic Duties? *Philosophy and Phenomenological Research* **60**: 591-606.
- Foley, Richard. 1985. What's Wrong with Reliabilism? *The Monist* **68**: 188-202.
- Goldman, Alvin. 1992. *Liaisons – Philosophy Meets the Cognitive and Social Sciences*. Cambridge, Mass.: MIT Press.
- Goldman, Alvin. 1986. *Epistemology and Cognition*. Cambridge, Mass.: Harvard University Press.
- Goldman, Alvin. 2011. O que é a crença justificada? Trad. Marques, L. H., S. R. N. Miranda e D. Murcho, URL: <http://criticanarede.com/justificacao.html>, acessado em 21 de fevereiro de 2017.

<sup>76</sup> Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio à pesquisa da qual o presente artigo é um resultado. Agradeço ainda aos úteis comentários de Waldomiro José da Silva Filho e Eros Carvalho e à revisão cuidadosa dos pareceristas do *Compêndio em Linha de Problemas de Filosofia Analítica*, Célia Teixeira e Domingos Faria, sobre versões anteriores deste artigo. Os erros e falhas que restam, claro, são de minha inteira responsabilidade.

- Greco, John. 2000. *Putting Skeptics in their Place – The Nature of Skeptical Arguments and their Role in Philosophical Inquiry*. Cambridge: Cambridge University Press.
- Greco, John. 2004. How to Preserve your Virtue while Losing your Perspective. In Greco, J. (ed.) *Ernest Sosa and his Critics*. Blackwell.
- Greco, John. 2010. *Achieving Knowledge – A Virtue Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- Greco, John e Turri, John. 2015. Virtue Epistemology. *The Stanford Encyclopedia of Philosophy*, E. Zalta (ed.), URL <http://plato.stanford.edu/archives/fall2015/entries/epistemology-virtue/>, acessado em 6 de maio de 2016.
- Haack, Susan. 1993. *Evidence and Inquiry - Towards Reconstruction in Epistemology*. Oxford: Blackwell.
- Henderson, David e Terence Horgan. 2000. Iceberg Epistemology. *Philosophy and Phenomenological Research*. **61**: 497-535.
- Henderson, David e Terence Horgan. 2011. *The Epistemological Spectrum – At the Interface of Cognitive Science and Conceptual Analysis*. Oxford University Press.
- Kahneman, Daniel. *Rápido e Devagar – Duas Formas de Pensar*. Trad. C. A. Leite. Rio de Janeiro: Objetiva.
- Kornblith, Hilary. 2002. *Knowledge and its Place in Nature*. Oxford University Press.
- Kornblith, Hilary. 2004. Sosa on Human and Animal Knowledge. In Greco, J. (ed.) *Ernest Sosa and his Critics*. Blackwell.
- Kornblith, Hilary. 2008. Knowledge Needs no Justification. In Smith, Q. (ed.) *Epistemology – New Essays*. Oxford: Oxford University Press.
- Lehrer, Keith e Stewart Cohen, S. 1983. Justification, Truth, and Knowledge. *Synthese* **55**: 191-207.
- Littlejohn, Clayton. The New Evil Demon Problem. *The Internet Encyclopedia of Philosophy*, URL <http://www.iep.utm.edu/evil-new>, acessado 7 de fevereiro de 2017.
- Mayr, Ernst. 1961. Cause and effect in biology. Kinds of causes, predictability and teleology are viewed by a practicing biologist. *Science* **134/3489**: 1501-1506.
- McGinn, Colin. 1999. *Knowledge and Reality – Selected Essays*. Oxford University Press.
- Millikan, Ruth. 1984. Naturalist Reflections on Knowledge. *Pacific Philosophical Quarterly* **65**: 315-334.
- Nozick, Robert. 1981. *Philosophical Explanations*. Oxford: Clarendon.
- Pritchard, Duncan. 2005. *Epistemic Luck*. Oxford: Oxford University Press.
- Proust, Joëlle. 2012. Metacognition and Mindreading: one or two Functions? In Beran, Michael et al. (eds.) *Foundations of Metacognition*. Oxford: Oxford University Press.
- Proust, Joëlle. 2013. *The Philosophy of Metacognition*. Oxford: Oxford University Press.
- Pylyshyn, Zenon. 2003. *Seeing and Visualizing – It's not What You Think*. Cambridge, Mass.: MIT Press.
- Rodrigues, Luis Estevinha. 2013. Conhecimento. In Branquinho, J. e R. Santos. (eds.) *Compêndio em Linha de Problemas de Filosofia Analítica*. Lisboa: Centro de Filosofia da Universidade de Lisboa. URL <http://compendioemlinha.letras.ulisboa.pt/conhecimento-luis-estevinha-rodriques/> Acessado em 12 de julho de 2016.
- Sosa, Ernest. 1991. *Knowledge in Perspective – Selected Essay in Epistemology*. Cambridge: Cambridge University Press.
- Sosa, Ernest. 2003. Beyond Internal Foundations and External Virtues. In Bonjour, L. e E. Sosa. *Epistemic Justification – Internalism vs. Externalism, Foundations vs. Virtues*. London: Blackwell.
- Sosa, Ernest. 2004. Replies. In Greco, J. (ed.) *Ernest Sosa and his Critics*. Blackwell.

- Sosa, Ernest. 2011. *Knowing Full Well*. Princeton University Press.
- Sosa, Ernest. 2013a. *Epistemologia da Virtude*, trad. L. P. Rouanet, São Paulo: Loyola [(2007) *A Virtue Epistemology – Apt Belief and Reflective Knowledge*, Oxford: Oxford University Press]
- Sosa, Ernest. 2013b. *Conhecimento Reflexivo*. Trad. C. C. Bartolotti, São Paulo: Loyola [2009. *Reflective Knowledge*. Oxford: Oxford University Press]
- Sosa, Ernest. 2015. Replies. *Analytica* **19**: 181-191.
- Stanovich, Keith E. 2011. *Rationality and the Reflective Mind*. Oxford: Oxford University Press.
- Turri, John. 2013. Bi-Level Virtuous Epistemology. In Turri, J. (ed.) *Virtuous Thoughts: the Philosophy of Ernest Sosa*. Dordrecht: Springer.
- Wedgwood, Ralph. 2002. Internalism Explained. *Philosophy and Phenomenological Research* **LXV**: 349-369.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.
- Williamson, Timothy. 2014. Knowledge First ; Knowledge Still First. In Steup, M., J. Turri e É. Sosa (eds.) *Contemporary Debates in Epistemology*. Wiley Blackwell.