

ENSAIOS EM FILOSOFIA DA MENTE

Ensaio Filosóficos



Maria Clara Dias

ENSAIOS EM FILOSOFIA DA MENTE

Maria Clara Dias

COPYRIGHT

© Maria Clara Dias

PRODUÇÃO GRÁFICA

Diogo Mochcovitch

REVISÃO

Diogo Mochcovitch

COLETÂNEA

ENSAIOS FILOSÓFICOS

CAPA

Pablo Picasso

ISBN

978-85-914916-5-0

©Maria Clara Dias, Rio de Janeiro, RJ, 2012.
Todos os direitos reservados.

INTRODUÇÃO

O QUE É FILOSOFIA?

A REFUTAÇÃO DO SOLIPSISMO E A NOÇÃO DE CRITÉRIO EM WITTGENSTEIN

CAUSAÇÃO INTENCIONAL

O FISCALISMO E O MISTÉRIO DA CONSCIÊNCIA FENOMENAL: OBSERVAÇÕES SOBRE A TEORIA PANIC DE MICHAEL TYE

O QUE PODEM SABER OS SERES NÃO-ONISCIENTES: SOBRE A RESPOSTA DE LYNNE BAKER AO FISCALISMO

QUINING THE GAP: A CONSCIÊNCIA FENOMENAL REVISADA*

TYING PHENOMENAL CONSCIOUSNESS AND QUINING THE GAP*

QUEM SOMOS NÓS? PRESSUPOSTOS E CONSEQUÊNCIAS DO PROGRAMA DE NATURALIZAÇÃO DO SELF

MIND AND PERSON IN A PHYSICAL WORLD

WHAT PERRY DOESN'T KNOW ABOUT MARY

REFERÊNCIAS BIBLIOGRÁFICAS

Introdução

Esta coletânea, assim como as duas anteriores, reuniu uma série de artigos que escrevi e publiquei, separadamente, ao longo dos anos. Sua reunião, em um único volume, visa fornecer aos interessados pela filosofia e, mais especificamente, pela filosofia da mente, um possível recorte temático e, quem sabe, uma iniciação a este fascinante universo através do qual buscamos lançar alguma luz sobre a nossa própria identidade, enquanto entidades às quais atribuímos, concomitantemente, propriedades físicas e mentais.

O primeiro capítulo versa sobre a própria atividade filosófica. "O que é Filosofia?" foi um artigo escrito para uma coletânea sobre o tema que idealizei durante o período em que dei aula na universidade de Ouro Preto. Para esta coletânea tive o privilégio de traduzir e publicar um artigo inédito de Ernst Tugendhat sobre a relação entre a Filosofia e as outras formas de saber. Meu próprio artigo está também baseado numa releitura do objeto da Filosofia realizada por Tugendhat em seu livro *Introdução à Filosofia analítica*. No meu artigo, defendo a tese de que a Filosofia se caracteriza por ser um domínio de regras ou conceitos universais, produzindo assim um saber ou um esclarecimento acerca das estruturas universais que regem tanto a linguagem ou o conhecimento humano, quanto o nosso agir.

O segundo artigo, "A refutação do solipsismo e a noção de critério em Wittgenstein", é o mais antigo da série. Nele procuro fornecer uma análise do argumento da linguagem privada como uma tentativa de refutação ao solipsismo/ceticismo. Como um sinal do que se revelou mais tarde o foco de meu interesse pela filosofia da mente, procuro analisar a crença no suposto caráter privado da linguagem acerca das sensações e estados psicológicos e apresentar seu desdobramento cético, através da assunção de uma premissa acerca da origem privada da linguagem. Procuro destacar a noção de critério como elemento central do argumento de Wittgenstein e justificar, através do caráter necessariamente objetivo/público dos mesmos, o caráter ao menos potencialmente público de todo e qualquer discurso significativo. Na mesma linha desta argumentação, foi escrito o último artigo desta coletânea, publicado mais de quinze anos depois.

O terceiro artigo foi escrito sob encomenda para um curso de introdução à Filosofia da Mente organizado pela Profa. Vera Vidal na Fundação Oswaldo Cruz. Trata-se, por isso, de um artigo mais descritivo, onde procuro apresentar a interpretação de Searle para o problema da relação causal entre o mental e o físico. O aspecto, no meu entender, mais positivo do artigo, está em esclarecer alguns conceitos centrais da filosofia da mente em geral e da perspectiva de Searle, em particular.

O quarto artigo fornece uma redescrição da problemática inicial, a saber, a preocupação com o caráter peculiar do mental, constitutivo de nossa subjetividade, no mundo físico, essencialmente público e/ou objetivo. "O fisicalismo e o mistério da

consciência fenomenal" foi publicado pela primeira vez, em 2005, no primeiro volume da revista *Abstracta: linguagem, mente e ação*. Nele apresento uma reconstrução da dificuldade apresentada para o fisicalismo pelos chamados objetos fenomenais; discuto as soluções analisadas por Tye e, finalmente, apresento uma resposta ao famoso "o argumento do conhecimento" à luz da teoria PANIC do caráter fenomenal.

O quinto capítulo foi escrito para um evento realizado pelo GT de Filosofia da Mente. Lynner Baker era a convidada especial e, para debater com ela escrevi "O que podem saber os seres não-oniscientes". O artigo foi publicado anos depois pela Revista Manuscrito (UNICAMP), v.xxv, p.89 - 104, 2003. Eu já havia discutido dois livros de Baker nos meus cursos de Filosofia da Mente. Embora, como o artigo deixe claro, discorde de sua posição, considero Baker uma grande filósofa e foi uma enorme satisfação descobrir que ela é também uma pessoa encantadora. Com Lynner Baker, pude vivenciar algo raro nos ambientes filosóficos: uma interlocutora respeitosa e atenta às nossas críticas. Meu artigo faz uma crítica ao teste dos contrafactuais proposto por Baker como critério para atribuição de poder causal a um determinado tipo de estado mental e à viabilidade aplicativa do mesmo. Procuro mostrar que o teste não responde ao problema da suficiência, ou seja, a possibilidade de identificarmos os contrafactuais corretos e suficientes, para a determinação da eficácia causal de eventos mentais.

O sexto artigo é uma triste e respeitosa homenagem a um filósofo e amigo querido. Ele foi encomendado, logo após o falecimento de Kaufmann, e deveria integrar um número especial da revista *Erudit* em sua homenagem. Sua versão em francês, Relier la Conscience phénoménale et éliminer le fossé explicatif, foi publicada em 2008, *Erudit*, v.35, p.513 - 524. Neste artigo, endosso a colocação do problema e o destaque dado ao conceito de consciência fenomenal propostos por Kaufmann, e corroboro a tese de que o problema do gap entre explicações mentais e fisicistas não pode ser encerrado como uma questão de princípio. Contudo, ao contrário de Kaufmann e todos os demais defensores dos *qualia*, defendo a tese de que as chamadas qualidades *sui generis* dos eventos mentais, reunidas sob o título de *qualia*, se reportam a eventos físicos com características peculiares. Estes constituiriam, tal como sugere Tye, o conteúdo representacional, não-conceitual, do *input* de processos físicos cujo *output* seriam outros estados intencionais de conteúdo proposicional, sendo estes últimos passíveis, ao menos *prima facie*, de uma explicação funcionalista.

O capítulo sete é apenas a versão em inglês do anterior.

O oitavo capítulo teve sua versão em inglês "Mind and Person in a Physical World" publicada no livro *Mente, linguagem e ação - textos para discussão*.1 ed. Porto:

Campo das Letras, 2009, v.1, p. 43-53. Sua versão em português compõe também a coletânea *Ensaio sobre a Moralidade* desta coleção. A razão é que neste artigo finalmente reuni as convicções adquiridas através do meu percurso pela Filosofia da Mente e minhas convicções acerca da natureza e do objeto da moralidade e da justiça. Desta forma, procuro apresentar um modelo do que seria uma concepção naturalista do ser humano e, em seguida, verificar até que ponto tal concepção é ou não capaz de resgatar de forma abrangente e coerente nossas diversas convicções morais. A versão em inglês constitui o nono capítulo desta série.

O décimo capítulo apresenta duas peculiaridades, foi escrito apenas em inglês e foi meu primeiro trabalho de coautoria. Por esta razão, optei por apresentá-lo por último, embora seu tema, como mencionei antes, remeta ao início desta série e a minhas primeiras investigações no campo da Filosofia da Mente. Este artigo decorre de um curso oferecido com o prof. Wilson Mendonça sobre um livro, então ainda inédito, de John Perry. Foi escrito para um congresso internacional de Filosofia da Mente em Lisboa. Em seguida, foi aceito e apresentado na Holanda, em um evento em homenagem a John Perry. "What Perry Doesn't Know About Mary" teve sua primeira versão impressa, anos depois, em *A Explicação da Interpretação Humana*. 1 ed. Lisboa : Edições Colibri, 2006, v.1, p. 359-368. O artigo, escrito com a colaboração de Wilson Mendonça, discuti a interpretação de Perry para o argumento do conhecimento, conhecido como "o caso de Mary" e conclui que a solução oferecida por Perry fracassa ao subsumir *knowing-how* a *knowing-that*. A leitura do livro de Perry, assim como o trabalho de elaboração do artigo, fizeram com que eu retomasse meus velhos tempos de leitura das *Investigações Filosóficas*.

Quem acompanha a leitura destas três introduções deve sair com a impressão que no fundo todos estes artigos são com o Bolero de Ravel, repetições sobre o mesmo tema. Os mais críticos talvez percebam tudo isso com enorme tédio. Os mais céticos, devem suspeitar de que no fundo forjei uma certa unidade temática. Em minha defesa posso apenas dizer que fui efetivamente surpreendida pela monotematicidade dos meus interesses filosóficos e que esta percepção, longe de me provocar a sensação de tédio, gerou em mim uma sensação de autenticidade e integridade na qual deposito grande parte do meu atual sentimento de autorrealização. É claro que a unidade e integridade de um pensamento não o torna necessariamente correto, aceitável, nem mesmo interessante. Mas esta etapa já foge ao meu controle. Caberá aos meus leitores julgar, se tive ou não algo a dizer. Da minha parte, sinto um misto da satisfação de quem consegue, enfim, ordenar a própria vida e de uma mãe que, finalmente, gera o filho durante anos desejado e, em seus mínimos detalhes, projetado. Se o filho será amado, isso nenhuma mãe pode garantir. Podemos apenas dar o melhor de nós e desejar que isso seja o suficiente para que aqueles que, com tanto cuidado e dedicação, ajudamos a construir obtenham também o acolhimento alheio.

O que é Filosofia?¹

Ser filósofo ou fazer filosofia costuma significar, na nossa vida cotidiana, ver ou perceber o mundo para além ou aquém da nossa experiência comum. O filósofo é muitas vezes o “metafísico”, ou seja, aquele cujo saber está além da experiência sensível, do aqui e agora das nossas determinações espaço-temporais. Mas o filósofo é também aquele que diz libertar a nossa compreensão do mundo e a nossa linguagem comum das garras da metafísica, ou seja, dos devaneios da razão especulativa que, incessantemente, nos lança para além dos limites do mundo inteligível. Quem é o filósofo afinal? É possível retecer o fio da chamada filosofia de forma a estabelecer um denominador comum ao que possamos considerar como sendo *a tarefa* da filosofia? Proponho aqui que partamos de um princípio bastante trivial: a Filosofia é uma prática, um método, um modo de proceder. Dito isto, não precisamos ainda nos posicionar com relação ao que possa, ou não, vir a ser o seu objeto. Resta, contudo, esclarecer em que esta prática, dita Filosofia, pode se distinguir de outras formas de proceder, tais como, por exemplo, de uma prática desportiva, científica ou artística.

Se pudermos aceitar que o que define práticas distintas são, antes de qualquer coisa, as regras empregadas em cada uma, então podemos concluir que devemos ser capazes de distinguir as regras de que nos servimos ao fazer Filosofia, das regras que empregamos na prática de um esporte, na investigação científica ou na arte. Sobre tal constatação não recai, no entanto, nenhum juízo de valor. Trata-se tão somente do reconhecimento de que as mesmas são práticas distintas. Nossa indagação acerca do que é Filosofia, pode ser então remontada nos seguintes termos: que características possuem as regras da prática dita filosófica? Para responder a tal questão pretendo eleger como estratégia a análise da caracterização do que seja Filosofia fornecida por Aristóteles e Kant. Em seguida, pretendo demonstrar como o projeto filosófico destes dois autores pode ser retomado pela chamada filosofia analítica da linguagem. Com isto, pretendo defender a tese de que a Filosofia se caracteriza por ser um domínio de regras ou conceitos universais, produzindo assim um saber ou um esclarecimento acerca das estruturas universais que regem tanto a linguagem ou o conhecimento humano, quanto o nosso agir.

Em que consiste a tarefa da Filosofia? Na *Metafísica*², Aristóteles fornece uma análise dos diversos modos de ser do ente e outorga à filosofia a tarefa de

¹ O presente artigo foi escrito para um seminário introdutório, acerca da questão “o que é Filosofia”, realizado na Universidade Federal de Ouro Preto e está baseado no livro de E. Tugendhat: *Einführung in die sprachanalytische Philosophie*, Frankfurt 1976. A rapidez com que é realizada a passagem de uma concepção de Filosofia a outra, o que certamente muito deixará a desejar ao leitor mais escrupuloso, deve ser compreendida como uma consequência inevitável do caráter introdutório do evento em questão.

investigar o seu modo de ser primeiro, ou seja, aquele a partir do qual todos os outros modos se dizem: sua substância ou essência. A substância de cada ente seria o denominador comum aos seus múltiplos modos de ser. Neste sentido, diríamos então que algo se caracteriza, enquanto um ente, por ser uma substância ou por estar relacionado à mesma, enquanto acidente, privação, corrupção, etc. Conhecer o "ser do ente", é conhecer a sua essência e os primeiros princípios de toda e qualquer entidade.

Enquanto unidade de referência entre os diferentes domínios e modos de ser das entidades, a essência ou substância não pode ser conhecida empiricamente, pois toda experiência é sempre experiência desse ou daquele domínio particular. Sendo assim, a ciência do ente, enquanto ente, só poderia assumir a forma de um conhecimento dos princípios relativamente aos quais dizemos que algo é um ente. Enquanto às ciências particulares caberia o estudo dos diferentes gêneros do ser, i.e., dos diferentes domínios da realidade; caberia à ciência do ente, enquanto ente, a investigação dos princípios comuns a todas as ciências. A filosofia seria, assim, o saber acerca dos primeiros princípios e estes são aqueles a partir dos quais tudo existe e tudo é conhecido. Ora, aquele que é primeiro na ordem do conhecimento deve ser imediatamente conhecido. A intuição seria, assim, o correlato cognitivo dos princípios. A filosofia passa a ser definida como um suposto conhecimento intuitivo de uma dimensão não-empírica da realidade, cujos princípios estariam na base de todo e qualquer conhecimento. A referência a essa suposta dimensão *a priori* da realidade explicaria o caráter universal que é atribuído ao saber filosófico.

Como podemos conhecer uma dimensão transcendente ou *a priori* da realidade? Em Aristóteles, os princípios, por serem a base de todo e qualquer conhecimento, devem ser imediatamente conhecidos, devem ser, portanto, intuídos. Pela intuição sensível, apreendemos uma determinada cor, uma determinada forma, esta ou aquela textura que nos permite identificar um objeto em particular. Como podemos compreender uma forma de intuição que nos forneça, não uma multiplicidade de dados sensíveis, mas princípios universais?

A partir da crítica à ideia de um saber intuitivo *a priori*, Kant³ redefine a filosofia como uma forma *a priori*, porém discursiva (ou não-intuitiva), de conhecimento. O conhecimento filosófico é o conhecimento racional através de conceitos. Neste sentido, Kant o distingue tanto do conhecimento empírico (aquele resultante de inferências a partir de experiências particulares), como do conhecimento dito matemático, ou seja, o conhecimento a partir da construção de conceitos. Mas como pode ser possível um conhecimento acerca da experiência que não esteja baseado em algum tipo de intuição? Ou seja, como são possíveis proposições filosóficas, ou em termos kantianos, juízos sintéticos *a priori*?

² Ver, Aristóteles, *Metafísica*.

³ Ver, I. Kant, *Kritik der reinen Vernunft*, Werke, Berlim 1902.

A solução kantiana consiste em demonstrar a validade de tais proposições como condição de possibilidade de toda e qualquer experiência. Conhecer *a priori* o "ser do ente" ou a "coisa em geral" consistirá, portanto, em conhecer as condições de possibilidade da experiência em geral.

Ora, só podemos conhecer *a priori* das coisas em geral, o que nós mesmos projetamos na sua natureza, afirma Kant. Neste sentido, caracterizaríamos a pluralidade dos gêneros e modos do que existe como entes, não porque esses seriam substâncias, ou acidentes de substâncias, ou privações e corrupções de substâncias, mas, simplesmente, porque podemos representar tudo o que há, como um objeto para a nossa consciência, abstração feita do modo diverso como ele exista. Deste modo, Kant circunscreve o projeto filosófico aos limites da própria consciência. Caberia à filosofia não mais a investigação das coisas tais como estas são em si mesmas, mas tal como podem vir a se constituir como um objeto para a nossa consciência.

Tendo em vista nosso próprio modo de apreender a realidade e refleti-la, a filosofia só pode ser concebida como um conhecimento *a priori* das condições segundo as quais algo pode vir a se tornar objeto do nosso conhecimento. Através da sensibilidade, apreendemos um múltiplo dado espaço-temporalmente, ou seja, o dado enquanto subsumido às formas da nossa sensibilidade, a saber: espaço e tempo. Através das chamadas categorias ou conceitos puros do entendimento, representaríamos o que é dado de modo desconexo como um objeto. A enticidade do ente ou a objetualidade das coisas não é determinada, portanto, por referência à essência, mas pela referência a uma consciência transcendental.

Para responder à possibilidade de falarmos - nos limites da consciência - de um conhecimento acerca da realidade, Kant irá demonstrar a unidade sintética do múltiplo dado como uma condição de possibilidade da própria unidade da consciência. De acordo com suas palavras:

*"(...) somente pelo fato de poder conceber numa consciência o múltiplo das representações, denomino todas de minhas representações, pois, do contrário, teria um eu mesmo tão multicolor e diverso quanto tenho representações das quais sou consciente"*⁴ (*Dedução Transcendental*)

ou ainda:

"(e)stou consciente de minha existência como determinada no tempo. Toda determinação temporal pressupõe algo permanente na percepção. Mas este permanente não pode ser algo em mim, pois precisamente minha existência no tempo pode ser, pela primeira vez, determinada por este permanente. Portanto, a percepção

⁴ Idem, ibidem, *Dedução Transcendental*.

deste permanente só é possível por uma coisa fora de mim e não pela representação de uma coisa fora de mim."⁵ (Refutação ao Idealismo)

Para provar a aplicação dos princípios que atestam validade ao nosso conhecimento acerca dos objetos, Kant assume como ponto de partida uma premissa cuja aceitação não possa ser colocada em dúvida, uma proposição autoverificável, ou seja, cuja tentativa de contradizê-la seja necessariamente falsa. Tal premissa é fornecida pelo próprio cético, como sendo a base a partir da qual foi possível colocar em questão o conhecimento do mundo externo, qual seja, o "eu penso". Se tivermos pensamentos que, por ora, se revelam de tal forma obscura que não podemos ao certo inferir-lhes uma causa no mundo externo, podemos então colocar em questão o caráter necessário da relação entre nossos pensamentos e o mundo externo e, por conseguinte, a própria possibilidade de um conhecimento objetivo. Não obstante, somos levados a aceitar como indubitável a certeza de que pensamos.

Se pudermos afirmar que temos pensamentos, então, devemos dispor de um saber acerca dos mesmos que nos permita compará-los e referi-los a um mesmo eu. Devemos ser capazes de nos autoatribuí-los. Em termos kantianos, esta capacidade de nos autoatribuir pensamentos significa o poder de reuni-los em uma autoconsciência ou ainda conformá-los às condições da unidade sintética da consciência (apercepção).

O fundamental da prova kantiana consiste, assim, em tentar mostrar a aplicação dos princípios que atestam a validade objetiva do nosso conhecimento acerca da experiência como condição de possibilidade da própria unidade sintética da consciência (apercepção), estabelecendo, assim, uma relação entre princípios. Relação esta entre a unidade sintética originária da consciência (apercepção) - condição de possibilidade do saber acerca de nossos pensamentos - e as condições de possibilidade da própria unidade sintética da consciência (apercepção), unidade sintética do múltiplo dado. Ao provar a necessária relação entre ambas as condições, Kant torna impossível a seu adversário aceitar a autoatribuição e, ao mesmo tempo, negar a validade objetiva de nosso conhecimento acerca de objetos. Estabelece, portanto, uma relação de pressuposição entre nosso saber acerca de nossos pensamentos e a aplicação de princípios que atestam objetividade.

O conhecimento filosófico, ou seja, o saber acerca dos princípios de todo e qualquer conhecimento possível é assim demonstrado com base na análise da consciência subjetiva que temos acerca de nossa própria experiência. Os principais passos da argumentação kantiana podem ser reconstruídos de acordo com as seguintes etapas:

1. Premissa factual autoverificável - "eu penso" - devo ser capaz de me autoatribuir experiências.

⁵ Idem, ibidem, Refutação ao Idealismo.

2. Regressão às condições de possibilidade da autoatribuição.

3. Dada a validade da autoatribuição, segue-se a validade das condições que a tornam possível. Prova da validade dos princípios que fornecem a garantia do nosso conhecimento acerca do mundo externo.

Conclusão: é lícito nos pronunciarmos acerca do mundo externo, em outros termos, nosso conhecimento acerca da experiência possui validade objetiva.

Ao admitirmos a capacidade de nos autoatribuirmos a consciência de dados sensíveis, estaríamos pressupondo a aplicação de certos princípios como condição de possibilidade da própria autoatribuição. Tais princípios se revelam, portanto, uma condição de possibilidade de todo e qualquer saber acerca da experiência.

Uma fundamentação semelhante a das chamadas proposições filosóficas podemos encontrar também na filosofia moral kantiana. Em sua *Crítica da Razão Prática*⁶, Kant procura fundamentar um princípio moral como condição de possibilidade da consciência do nosso próprio agir: todo aquele que tem consciência de seu próprio agir, age refletidamente, ou seja, age racionalmente e deve, portanto, ser capaz de justificar suas próprias ações.

Kant distingue a justificativa do agir moral em dois níveis: (i) ações são justificadas a partir de princípios subjetivos ou máximas; (ii) máximas devem ser fundamentadas a partir de um princípio superior que, por sua vez, independe das inclinações subjetivas de cada um. Tal princípio deve consistir na condição formal da racionalidade do agir. Esta condição é, em Kant, o princípio de universalização das máximas.

Em resumo, poderíamos dizer:

(1) Ter consciência do agir significa: (i) reconhecer uma relação entre o agir e as máximas do mesmo; (ii) reconhecer uma relação entre as máximas do agir e um princípio que as justifique.

(2) Este princípio só pode ser um princípio formal, ou seja, a própria condição da racionalidade do agir, ou seja, o princípio de universalização das máximas.

(3) Agir moralmente significa agir com base em máximas que possam ser justificadas a partir do princípio de universalização.

Conclusão: todo aquele que tem consciência de suas próprias ações deve reconhecer a validade da lei moral como condição de possibilidade desta consciência, ou seja, aceitar (1), (2) e (3).

De acordo com esta interpretação, há, portanto, em Kant, uma característica comum às chamadas proposições filosóficas, portanto, um mesmo método de

⁶ I. Kant, *Kritik der praktischen Vernunft*, Werke, Berlim 1902.

comprovação de sua pretensão de validade: trata-se de demonstrá-las como constituindo uma condição de possibilidade da consciência de nossa própria experiência (no caso da filosofia teórica) ou da consciência de nossas próprias ações (no caso da filosofia prática).

Podemos agora perguntar: como podemos identificar ou discriminar modos diversos de consciência? Se rejeitarmos a ideia de uma intuição de essências ou de uma intuição interna, algo assim como um “olhar interiormente”, só poderemos esclarecer nossos distintos estados de consciência a partir de uma análise da própria linguagem. Mas será que, com isto, não deixaremos escapar algo de fundamental? A consciência, enquanto consciência de objetos, ou seja, enquanto voltada ou direcionada para algo, é também chamada consciência intencional.⁷ Para justificar a pretensão inicial desta exposição e mostrar como é possível, a partir da filosofia analítica da linguagem, recuperar a questão fundamental da chamada filosofia transcendental, qual seja, a questão da possibilidade de um conhecimento acerca dos objetos, pretendo agora demonstrar que todos os modos de consciência intencional são, ou implicam, modos de consciência proposicional.

São modos de consciência intencional o perceber, o imaginar, o afirmar, o desejar, o temer, o amar etc. Todos esses verbos são chamados transitivos e se caracterizam por possuir um objeto gramatical. A intencionalidade é, assim, uma relação. Resta, contudo, analisar em que essa relação se distingue das demais ou como tais verbos se distinguem de verbos transitivos não intencionais. Passemos, então, à análise dos objetos capazes de preencher a função de objeto gramatical em enunciados intencionais.

Entre os verbos intencionais, podemos discriminar (1) aqueles cujo complemento é um termo singular que representa um objeto concreto espaço-temporal como, por exemplo, “Amar Pedro”; (2) aqueles que exigem como complemento sentenças nominalizadas, ou seja, expressões linguísticas da forma “este p”, em que p representa um estado de coisas como, por exemplo, “Pensar que está chovendo lá fora” e, finalmente, (3) verbos que podem ter como complemento tanto termos singulares quanto sentenças nominalizadas como, por exemplo, “Temer a chuva” e “Temer que lá fora esteja chovendo”. Esses três tipos de verbos irão caracterizar dois modos de consciência intencional: uma consciência intencional não proposicional e uma consciência intencional proposicional.

Se houvesse apenas um modo de consciência intencional: o proposicional, então seria possível apresentar um critério de distinção entre as relações intencionais e as demais: só na intencionalidade pode haver uma relação entre um objeto concreto,

⁷ Acerca do conceito de intencionalidade ver: F. Brentano, *Psychologie vom empirischen Standpunkt*, Leipzig 1924/25 e E. Husserl, *Logische Untersuchungen*, Halle 1900/01.

qual seja, uma pessoa, e uma proposição. Havendo dois modos de consciência intencional, resta-nos três alternativas: (1) ou bem conseguimos encontrar alguma característica comum, (2) ou reconhecemos que entre esses dois modos de consciência não há nada em comum (o que tornaria o próprio conceito de intencionalidade um pseudo conceito); (3) ou conseguimos demonstrar que toda forma de consciência intencional é, ou implica, um modo de consciência proposicional. Elegerei aqui a terceira alternativa.

De acordo com Brentano⁸, o que distingue uma relação intencional das demais é o fato de que o segundo termo dessa relação não precisa existir. Relações não intencionais, por sua vez, só são possíveis quando ambas as partes existem. Como podemos explicar essa peculiaridade das relações intencionais? Para que uma criança tema fantasmas, não é necessário que fantasmas existam, mas podemos dizer que ela deve, ao menos, acreditar que os mesmos existam. Ora, acreditar que algo exista, ou melhor, que haja um objeto ao qual pertença as propriedades implicadas pelo termo singular “fantasma”, é um modo proposicional de consciência. Do mesmo modo, quando digo que “vejo uma árvore” estou, ao mesmo tempo, afirmando (1) que há algo que possa ser identificado como uma árvore; (2) que aqui, no meu campo óptico, haja alguma coisa e (3) que o objeto em questão seja uma árvore. Assim sendo, meu enunciado será falso em três situações: (1) quando não houver nada que possa ser identificado como uma árvore, ou seja, quando não existirem árvores; (2) quando não houver nada no meu campo óptico ou ainda (3) quando aquilo que vejo não for uma árvore. Mais uma vez a análise de um modo de consciência intencional não-proposicional terá, portanto, revelado como pressuposto um modo de consciência intencional proposicional, qual seja, a crença de que aquilo a que nos referimos exista.

Mas o que acontece quando falamos ficticiamente de algo, ou seja, quando temos consciência de que o objeto ao qual nos referimos não existe? Estaríamos, então, como diria Wittgenstein, ingressando em um novo jogo de linguagem, algo como o jogo do “Faz de conta”. Diríamos agora: “Imagine que ...” e mais uma vez chegaríamos a um modo de consciência intencional proposicional. Podemos, portanto, concluir que a qualidade peculiar da chamada consciência intencional consiste na relação entre uma pessoa e uma proposição. Compreender os diversos modos de consciência significa, assim, compreensão dos diversos empregos de proposições na nossa linguagem.

A investigação acerca daquilo que confere unidade aos múltiplos modos de ser do ente (a questão tradicional acerca do ente enquanto ente) ou à multiplicidade de sensações ou dados sensíveis (a questão acerca da objetualidade das coisas em geral) pode ser, portanto, recolocada e resgatada como a investigação e análise

⁸ Ver, F. Brentano, *Psychologie vom empirischen Standpunkt*, Leipzig 1924/25

linguística do modo pelo qual nos referimos a objetos, nos diferentes domínios da realidade, a saber: o emprego de predicados e termos singulares em proposições predicativas.

O significado de uma expressão linguística em geral, e de um termo geral em particular, consiste no seu modo de uso. Neste sentido, compreender uma proposição predicativa elementar da forma “Fa” significa saber verificar publicamente se o objeto identificado pelo termo singular “a” é ou não o caso do predicado “F” em questão. Indicar o modo como um termo “F” se aplica a um objeto “a” é o que fazemos quando, por exemplo, explicamos o emprego do termo geral “vermelho” com base em exemplos positivos e negativos.

No que diz respeito à tarefa da Filosofia, podemos, agora, acrescentar: cabe ao filósofo fornecer uma elucidação dos nossos diversos modos de uso da linguagem e das pretensões por eles erguidas. Assim sendo, caberá ao filósofo a elucidação das regras constituintes dos diversos discursos que compõem a nossa linguagem, dentre estes os discursos científico, moral e artístico. Resta, contudo, lembrar que a antes mencionada *nossa linguagem* não deve ser compreendida como *a linguagem*, ou seja, como uma linguagem acabada e inequívoca. O que queremos compreender como sendo *a nossa linguagem* é muito bem ilustrado, como diria Wittgenstein⁹, pela metáfora de uma cidade. Colocamos uma rua aqui e outra ali e uma cidade se constrói, sem que pareça fazer sentido nos perguntarmos por onde começou e onde terminará.

⁹ Ver, L. Wittgenstein, *Philosophische Untersuchungen*, em: Schriften, Bd. 1.

A refutação do solipsismo e a noção de critério em Wittgenstein

O objetivo deste artigo é fornecer uma análise do argumento da linguagem privada como uma tentativa de refutação ao solipsismo/ceticismo.¹⁰ Para tal pretendo em primeiro lugar apontar para o surgimento da questão solipsista a partir da crença no caráter privado de nossa linguagem de sensações e estados psicológicos, e apresentar seu desdobramento cético a partir da aceitação de uma origem privada da linguagem. Em seguida, pretendo discutir duas interpretações do argumento da linguagem privada, Strawson e Bennett, e através das mesmas apontar para o que deve constituir o elemento central do argumento de Wittgenstein, a saber: a noção de critério. Para concluir, pretendo fornecer uma breve reconstrução do argumento da linguagem privada, entendido como uma tentativa de justificar o caráter, ao menos potencialmente, público de todo e qualquer discurso significativo.

I

Na nossa vida cotidiana estamos familiarizados com frases do tipo: "Só eu sei a minha dor", ou ainda, "meus mais profundos sentimentos ninguém jamais poderá compreender". Tais enunciados ilustram a crença de que nossas sensações e estados psicológicos são, no que lhes é essencial, inacessíveis a outros sujeitos. Acreditamos, assim, em uma linguagem cujo sentido é privilégio apenas do próprio falante, uma linguagem privada.

A aceitação de uma linguagem privada repousa na crença de que nossas sensações e estados psicológicos possuem tal natureza que apenas aquele que os vivencia pode compreendê-los. Instaure-se, portanto, um abismo entre a linguagem pública, na qual nos servimos de termos comuns para nomeá-los, e o sentido privado deste âmbito de experiências, a que apenas cada sujeito, privadamente, poderia ter acesso. O verdadeiro sentido destas palavras passa a manter uma tênue relação com nossa linguagem comum. Supomos compreender o uso de nomes para sensações, com base em nossos próprios estados, porém neste percurso escapa o que nos parece essencial, qual seja, a vivência. Aquilo que em cada um de nós fornece a garantia de que 'realmente' compreendemos estas palavras não pode ser transposto a outro, deste modo, somos levados a apenas acreditar que estejamos de fato nos entendendo. Confiamos que estejamos diante de outros sujeitos e, finalmente, transpomos nossa compreensão, acerca de nós mesmos, para os que se comportam de modo análogo.

¹⁰ Uma análise mais extensa desta questão pode ser encontrada em M.C. Dias, *O Argumento da Linguagem Privada: investigações filosóficas acerca do discurso significativo*, PUC/RJ, Rio de Janeiro 1989.

Ora, será que devemos aceitar, apenas a título de crença, a possibilidade de uma linguagem comum acerca das sensações e estados psicológicos? Se não temos acesso àquilo de que provém o sentido de tais termos quando empregados por outros sujeitos, então é possível que toda a pretensão de comunicarmos tais experiências não passe de uma mera ilusão, e que nos enganemos sempre que, por analogia, supusermos compreender um outro sujeito. Na medida em que supomos, portanto, que só podemos saber acerca dos estados alheios através do conhecimento de nossas próprias vivências, dos termos e comportamentos a elas por nós associados, tão tênues serão as razões pelas quais atribuiremos vivências comuns a terceiros, que já não distantes estaríamos, até mesmo, da possibilidade de duvidar que possamos crer na existência de outros sujeitos, enquanto seres dotados de predicados psicológicos.

Tomando este mesmo ponto de partida, podemos ainda chegar a uma nova acepção do que seja uma linguagem privada, desta vez não mais restrita ao âmbito das sensações e estados psicológicos, mas voltada para todo discurso. Para tal, basta que suponhamos, como parece ter sido o caso na modernidade, que todo o nosso discurso deva ser mediado pela consciência de nossas próprias impressões ou estados subjetivos, possuindo assim um caráter originalmente privado. Neste caso, poderemos estender a dúvida acerca de estados alheios à dúvida que concerne à própria possibilidade de um discurso objetivo.

Por certo, entre estas duas definições de linguagem privada, a primeira restrita ao âmbito das sensações e estados psicológicos e a última voltada para uma origem privada da linguagem, há uma progressão não necessária. Isto significa que podemos encontrar defensores da primeira que recusem a acepção mais forte, contudo não o oposto. Através da análise do argumento da linguagem privada, pretendo mostrar que a refutação à possibilidade de uma linguagem privada, ainda que compreendida apenas em sua primeira acepção, culmina igualmente na refutação à crença de uma origem privada da linguagem. Entre ambas, encontramos o mesmo pressuposto, o hiato entre nosso discurso comum e aquilo que garante sua validade, quer para o âmbito das experiências subjetivas, quer para o discurso acerca do mundo externo.

II

De acordo com a interpretação apresentada por Strawson em seu artigo “Exposition and criticism of Wittgenstein’s Investigations”¹¹, o argumento de Wittgenstein oscila entre duas teses:

“A tese mais forte diz que nenhuma palavra nomeia sensações (ou experiências privadas); e em particular a palavra “dor” não o faz (I.F. 293). A tese

¹¹ Strawson, P.F., *The Private Language Argument*, Macmillan St. Martins Press, 1971.

mais fraca fala que certas condições devem ser satisfeitas para a existência de uma linguagem comum na qual sensações são atribuídas àqueles que as têm.”¹²

Para a primeira tese, destinada ao ataque a uma linguagem privada, seriam apresentados dois argumentos. O primeiro argumento apontaria para a ausência de critérios de correção numa linguagem privada, com base na impossibilidade de que um recurso à memória¹³ pudesse fornecer tais critérios. Contudo, assinala Strawson, sendo este o ponto central do argumento, não haveria porque estabelecer o privilégio concedido à linguagem que se refere às sensações, pois, qualquer linguagem usada por apenas uma pessoa seria absurda.¹⁴ Ora, um observador, capaz de correlacionar o uso das expressões com as ações de um falante, pode vir assim a estabelecer uma regularidade que possibilite a determinação do significado da expressão. Define, portanto, critérios para verificar o uso correto de tal termo. O significado seria determinado por uma prática de uso, e o observador seria igualmente capaz de checar sua memória de forma a relacionar usos passados e presentes. Qual seria a especificidade, concedida às sensações, que tornaria impossível tal procedimento? O que impediria a checagem na linguagem privada das sensações seria a ausência de elementos manifestos no uso das expressões, impossibilitando assim o estabelecimento da relação entre o nome e a sensação.

O segundo argumento seria uma variante do primeiro, baseando-se no caráter expressivo das proposições do tipo “eu tenho dor”: [1] Em primeiro lugar, afirma-se que a expressão da dúvida não tem lugar em proposições deste tipo; [2] em segundo, segue-se que quando alguém diz ter dor, não identifica sua sensação com critérios, mas apenas a expressa. O termo “dor” não seria, então, o nome ou a descrição de uma sensação, mas funcionaria como um critério para que um terceiro pudesse atribuir uma sensação ao falante.

A segunda tese resulta da introdução de uma premissa acerca da linguagem: “Tudo que há para ser dito acerca do significado descritivo da palavra é dito quando é indicado qual critério as pessoas podem usar para empregá-la ou para decidir quando ela é ou não empregada corretamente.”

De acordo com Strawson, ao associar as duas teses, Wittgenstein cometeria o erro de supor impossível reconhecer ou identificar algo sem o uso de critérios, e, neste caso, impossível reconhecer e identificar sensações. Para Strawson, porém, uma sensação não é identificada por critérios, mas pode ser identificada por aquele que a vivencia. A identificação ou o reconhecimento da sensação só seria colocado quando

¹² Idem, ibidem, pp.27. Tradução livre.

¹³ O suposto recurso de Wittgenstein à impossibilidade de nos basearmos na memória será discutido na etapa seguinte, destinada à interpretação de Bennett.

¹⁴ Ver, idem, ibidem, pp.28.

o que estivesse em questão fosse a possibilidade de uma linguagem comum, impessoal, acerca das sensações. Em suas próprias palavras:

*“(...) é claro que não se utiliza critérios de identificação para o gosto. Se a questão ‘Qual é o critério de identificação aqui?’ é colocada, pode-se apenas responder: ‘Bem, o gosto ele mesmo’ (‘a sensação ela mesma’). É claro que a frase pela qual nos referimos a tais gostos envolve alusões ao que pode ser visto e tocado; a fim de que falemos uma linguagem comum.”*¹⁵

Deste modo, apenas o uso público de nomes para sensações requer critérios, e estes não são fornecidos com referência à sensação, mas pela associação a elementos observáveis. Esta posição faz com que Strawson recuse a primeira tese imputada a Wittgenstein, mas possa aceitar a segunda. Para concluir, Strawson procura assinalar qual teria sido o “equivoco” de Wittgenstein que o conduz à primeira tese, e propõe uma solução para a possibilidade de uma linguagem pública acerca das sensações. Em suas palavras:

*“Nesta possibilidade de um acordo geral nos julgamentos está a possibilidade de uma linguagem comum impessoal para descrever o que vemos, ouvimos e tocamos (cf. I.F.. 242). **But there is no such general agreement as to whether or not “it’s painful here”, as to what it feels like (as we misleadingly say) within.** Mas não há um tal acordo geral a respeito de ‘estar ou não dolorido aqui’, assim como acerca de como sentimos interiormente (tal como equivocadamente dizemos). Na ausência de um acordo geral no julgamento, é impossível uma linguagem comum; e aí está porque é impossível uma linguagem comum impessoal da dor.”*¹⁶

A solução seria, então, fazer da nossa própria “linguagem para sensações” uma linguagem para atribuir sensações a outros indivíduos, e faríamos isto com base no que vemos, ouvimos, e em tudo mais acerca de que seja possível um acordo. Com isso, o comportamento da dor passa então a servir de critério para que possamos atribuir dor a quem a expressa.

Pretendo agora analisar a plausibilidade da reconstrução proposta por Strawson como uma interpretação de Wittgenstein, e, em seguida, sua contribuição à polêmica acerca da possibilidade de uma linguagem privada.

A primeira tese imputada a Wittgenstein, qual seja, “nenhuma palavra nomeia uma sensação” traz como referência o parágrafo 293, um dos mais polêmicos das *Investigações Filosóficas*. Neste parágrafo, Wittgenstein nos fornece o exemplo da caixa com o besouro e conclui com os seguintes dizeres: “(...) quando se constrói a

¹⁵ Idem, ibidem, pp. 30. Tradução livre.

¹⁶ Idem, ibidem, pp.32. Tradução livre.

gramática da expressão da sensação segundo o modelo de ‘objeto e designação’, então o objeto cai fora de consideração, como irrelevante.”¹⁷.

Portanto, quando supomos que o significado de nossas palavras seja aquilo por elas designado, e este referente é privado, jamais poderemos compreender o uso de tal termo por outro, ou melhor, poderemos apenas acreditar compreender, e para tal não nos basearemos no objeto referido, já que a este jamais teremos acesso, mas no comportamento e situação de uso da expressão, através de uma analogia com o nosso próprio uso e suas manifestações.

A primeira vista, não dispomos de razão alguma para supor que, nesta passagem, esteja sendo dito que não nomeamos sensações, ao contrário, podemos supor que esteja apenas mais uma vez sendo ilustrada a crítica a uma concepção segundo a qual o significado de uma palavra é o seu referente. Associar esta crítica à primeira tese seria supor que dar um nome a uma sensação seja simplesmente associar um signo a um objeto, sem que se requeira para tal um contexto de uso da linguagem, onde uma série de outros elementos estão envolvidos e contribuindo de forma fundamental para a determinação do significado, suposição esta que Wittgenstein se empenha em recusar. Portanto, se o que Strawson pretende dizer é o que Wittgenstein rejeita - e neste caso, não que as sensações possam ser nomeadas, mas que para tal nada seja necessário além da própria sensação, - a interpretação parece correta, apesar de ambígua.

No decurso de sua exposição, particularmente com a introdução da segunda tese, Strawson torna mais claro o que pretende dizer com a primeira tese. Se com esta tese Wittgenstein rejeitasse a possibilidade de fornecermos nomes para sensações - tese que contraria o parágrafo 244, cujo objetivo é exatamente elucidar o modo como aprendemos a nomear sensações, - seria contraditório afirmar pela segunda tese a possibilidade de uma linguagem comum acerca das sensações. Para que a interpretação de Strawson seja coerente, devemos então reformulá-la nos seguintes dizeres: [1] não é possível dar nomes a sensações em uma linguagem privada e [2] certas condições devem ser satisfeitas para que seja possível uma linguagem pública acerca das sensações.

Para Strawson, no entanto, a base, e também o equívoco, da primeira tese está na suposição de que não podemos identificar ou reconhecer sensações privadamente, pois não disporíamos de qualquer critério de distinção. Somente a linguagem pública nos forneceria, assim, os critérios necessários para aplicar um nome a uma sensação. Concordo que este seja para Wittgenstein o cerne da questão, mas não faz o menor sentido supor que este seja também seu equívoco, a menos que estejamos dispostos a

¹⁷ Wittgenstein, L., *Philosophische Untersuchungen*, Suhrkamp, Frankfurt, 1984,293. Esta e as demais traduções de Wittgenstein foram retiradas da edição brasileira da série Pensadores, Abril Cultural, São Paulo.

reconhecer como equivocada a própria tentativa de fornecer um argumento contra a possibilidade de uma linguagem privada.

Em suma, a tese de Strawson segundo a qual é possível identificar e reconhecer uma sensação privadamente, sem qualquer critério, apenas pela própria sensação, é a tese defendida pelos que postulam a possibilidade de uma linguagem privada. Sua solução para o problema da possibilidade de uma linguagem pública acerca das sensações em nada se distingue da solução apresentada pelos céticos. A linguagem pública torna-se assim uma derivação da nossa linguagem privada. Identificamos nossas sensações e, em seguida, as associamos a elementos manifestos, de tal forma que, ao reconhecermos estes mesmos elementos em alguém, passamos a atribuir-lhe uma sensação determinada. Voltamos, assim, ao nosso ponto de partida, qual seja, podemos apenas acreditar que outras pessoas possuam sensações. Sem que sejamos capazes de mostrar a impossibilidade de identificarmos ou reconhecermos sensações, apenas com base em nossos estados internos ou na nossa própria sensação, jamais teremos eliminado a premissa fundamental dos defensores de uma linguagem privada.

III

Enquanto Wittgenstein procura fornecer um argumento contra a possibilidade de uma linguagem privada, Strawson formula um argumento que justifique a possibilidade de uma linguagem pública. Qual seria a posição de Bennett diante destas duas perspectivas?

Em Bennett, o polêmico argumento da linguagem privada é discutido em um capítulo destinado à reconstrução da “refutação ao idealismo”¹⁸. Wittgenstein é introduzido no contexto da discussão acerca das experiências internas e da confiabilidade da memória. De acordo com Bennett, uma pessoa que se mantivesse restrita ao âmbito de suas experiências internas jamais poderia discorrer acerca do passado, pois neste caso, todos os seus estados passados estariam reduzidos a recordações presentes, ou ainda, o próprio passado estaria reduzido ao presente. Tornar-se-ia, assim inútil qualquer distinção entre ser e recordar, simplesmente porque o próprio conceito de passado já não faria sentido. Em sua interpretação, a opinião de Kant e Wittgenstein acerca das experiências internas seria, contudo, bem mais radical:

“Argumentei que uma experiência interna que não fosse então exteriorizada não poderia ser subsumida ao ‘conceito de passado’, enquanto Kant pensa que não poderia ser subsumida a ‘qualquer conceito’, e seria, por conseguinte, impossível. Similarmente, Wittgenstein parece concluir não que uma linguagem puramente

¹⁸ Bennett, J., *Kant's Analytic*, Cambridge University Press, Cambridge 1966.

privada não possa ter um tempo passado, mas que tal linguagem não poderia existir.”¹⁹

Entre uma posição que sustenta apenas que algo não possa ser submetido ao conceito de passado, e outra que afirma sua impossibilidade, há, sem dúvida, uma grande distância. Para sustentar a segunda posição, dois argumentos tornaram-se clássicos: a refutação ao idealismo e o argumento da linguagem privada. Em que medida ambos teriam alcançado êxito é o que Bennett pretende investigar.

Bennett defende que alguém possa ter conhecimento de experiências internas presentes sem que para tal tenha que recorrer a experiências externas. No entanto, admite que, para podermos saber acerca das experiências internas de um terceiro, seja necessário sermos capazes de distinguir (i) aquilo podemos saber acerca do outro (ii) do que podemos saber acerca de nós mesmos. Contudo, o conhecimento que cada qual possui de si próprio não pode decorrer apenas de experiências internas presentes. Nossa “autoconsciência” só pode ser expressa pela capacidade de correlacionarmos enunciados do tipo “assim eu sou agora” com enunciados envolvendo experiências passadas, “assim eu era”. Sem este recurso às experiências passadas, não seria satisfeita a condição segundo a qual nos tornamos capazes de reconhecer experiências alheias, qual seja, a autoconsciência ou a capacidade de correlacionar experiências diversas, e atribuí-las a um mesmo sujeito. Na interpretação de Bennett, este seria o procedimento de Kant na “Refutação ao idealismo”, e o que tornaria justificada sua pretensão. Restaria, assim, avaliar o argumento fornecido por Wittgenstein.

A primeira formulação do argumento da linguagem privada tomaria como prioridade a noção de “privado de fato”. Nesta perspectiva, estaria sendo afirmado: [1] só podemos confiar em nossa memória na medida em que esta se apoia em comprovações, e [2] as comprovações necessárias devem ser fornecidas por estados objetivos ou impessoais. Bennett critica esta formulação, pois, a aceitação de [1] faria com que nenhuma fonte de conhecimento do passado fosse digna de confiança e, como consequência, o recurso a [2] seria inútil. Wittgenstein daria respaldo a esta interpretação ao afirmar a impossibilidade de obedecermos a uma regra privadamente.

Na segunda formulação, o argumento estaria centrado nas condições segundo as quais podemos afirmar estados internos em uma linguagem pública. Na opinião de Bennett, este argumento teria êxito total, porém, nenhuma relação com a pretensão de mostrar a impossibilidade de uma linguagem necessariamente privada. Aqui, Wittgenstein estaria apenas mostrando a relação entre as causas e a conduta de alguém diante de uma sensação, e o significado do que é dito publicamente acerca da mesma. Causas e comportamentos seriam elementos indispensáveis ao informe público de nossos estados internos. Deste modo, o que o argumento da linguagem

¹⁹ Idem, *ibidem*, pp.209. Tradução livre.

privada acabaria por mostrar seria bem menos do que o pretendido. Com isso, Bennett conclui:

“Se Wittgenstein pudesse mostrar que [i] uma regra não pode ser obedecida privadamente, seguir-se-ia que [ii] não pode haver uma linguagem que seja de fato privada, enquanto talvez pudesse se seguir que [iii] alguém que possa subsumir seus estados internos a uma linguagem deva, portanto, de alguma maneira, engajar-se em comportamentos físicos, e/ou estar sujeito a causas físicas. A aceitação de [iii] pode predispor alguém a pensar que [iv] a concomitância causal e comportamental de nossos relatos públicos de nossos estados internos é de algum modo vinculada com o significado de tais relatos. Eu tentei no entanto mostrar que Wittgenstein tem um excelente argumento a favor de [iv] o qual não requer o nebuloso movimento (cloudy move) de [iii] a [iv] ou a premissa inicial [i], para a qual ele não tem de todo fundamentos.”²⁰

Podemos concordar com Bennett que tanto Kant quanto Wittgenstein tenham tentado fornecer argumentos em favor de uma impossibilidade. Uma impossibilidade que em Kant envolve um ‘interno’ e em Wittgenstein um ‘privado’. Tratar-se-ia, em Kant, de uma experiência subjetiva capaz de ser discriminada sem qualquer recurso a uma experiência objetivo. Segundo Bennett, tal experiência seria impossível porque sua condição, a autoconsciência, pressuporia aquilo que ela própria deveria excluir.²¹ Resta, contudo, uma questão: por que devemos poder correlacionar experiências presentes e experiências passadas para dispormos de um saber acerca de nós mesmos? Eis o ponto central que Bennett parece negligenciar.

O que faz, em Kant, com que não possamos compreender uma experiência interna, que não esteja correlacionada a experiências externas, ou melhor, objetivas, e ao tempo, é simplesmente o fato de que estes últimos são os critérios dos quais dispomos para discriminar nossas experiências em geral. Portanto, sem uma referência a experiências objetivas e ao tempo, não haveria qualquer critério de discriminação, e acerca de uma experiência indiferenciada não haveria saber ou discurso possível. Deste modo, o que Kant recusa não é a simples possibilidade de experiências internas, mas sim o saber acerca de uma experiência indiferenciada, caótica. Para tal, deverá mostrar que ‘critérios internos’ não produzem qualquer distinção, ou melhor, não podem funcionar com critério.

Vejamos agora em que constituiria o ‘privado’ em Wittgenstein. Começarei pelo argumento que, segundo Bennett, concerne ao ‘privado de fato’, baseado na

²⁰ Idem, *ibidem*, pp.214. Tradução livre.

²¹ Em outro lugar (“Kant: os limites da razão crítica”, *Síntese Nova Fase* 72, 1996) procurei mostrar que a determinação temporal é a condição de possibilidade de nos autoatribuirmos experiências, isto, porém, porque ela exprime critérios de discriminação e identificação, sem os quais jamais seríamos capazes de correlacionar experiências e reuni-las em uma unidade.

impossibilidade de confiarmos na memória. Que passagem de Wittgenstein autoriza tal interpretação? A citação de Bennett é o parágrafo 265 das *Investigações Filosóficas*:

“Imaginemos uma tabela que exista apenas em nossa imaginação; algo como um dicionário. Por meio de um dicionário podemos justificar a tradução da palavra x para a palavra y. Mas devemos chamar a isto também de justificação, se esta tabela é consultada apenas na imaginação? - ‘Ora, trata-se então de uma justificação subjetiva’ - Mas a justificação consiste em que se apele a uma instância independente. - ‘Mas posso também apelar para uma recordação partindo de outra. Não sei (por exemplo) se guardei corretamente o horário da partida do trem e, para o controle, evoco a imagem da página do livro de horário. Não temos aqui o mesmo caso?’ - Não, pois esse processo deve verdadeiramente provocar a recordação ‘correta’. Se não precisasse ‘comprovar’ a exatidão da imagem mental do livro de horários, como poderia esta comprovar a exatidão da primeira recordação? (Como se alguém comprasse vários exemplares do jornal do dia, para certificar-se de que ele escreve a verdade). Consultar uma tabela na imaginação é tão pouco consultar uma tabela, quanto a representação do resultado de um experimento imaginado é o resultado do experimento.”²²

Ora, parece claro que o que Wittgenstein está colocando em questão não é a memória, como supõe Bennett, mas o recurso ao ‘mesmo’ como critério de correção. E, sob este aspecto, vale como exemplo tanto aquele que recorre à memória para recordar algo esquecido, como o que compra vários exemplares de um mesmo jornal para verificar a correção de um relato. O que está sendo criticado é a concepção de uma experiência interna, cujos critérios de discriminação fossem igualmente internos, ou seja, dependentes da mesma. Neste caso, o próprio critério se confundiria com aquilo a que cabe discriminar, e restaria apenas uma experiência indiferenciada. Por conseguinte, o que torna a linguagem privada impossível é a ausência de critérios. Graças a este mesmo fato, podemos dizer agora que seguir regras privadamente, ou dispor de qualquer saber acerca de uma experiência indiferenciada, seja igualmente impossível. Deste modo uma experiência interna, assim como uma linguagem privada, é aquela na qual não há discriminação possível, na qual faltam critérios.

Nada em Wittgenstein parece impedir que confiemos em nossa memória, assim como nada nos indica que nossa memória seja indiferenciada ou caótica. Se uma linguagem ‘privada de fato’ é aquela na qual alguém se encontra sozinho seguindo regras, ela em nada pode nos parecer problemática, pois, será então potencialmente comunicável - isto significa que deverá possuir critérios de correção. Quando Bennett nos fala na recusa de uma linguagem necessariamente privada, talvez esteja supondo que pudesse haver uma linguagem privada, porém,

²² Wittgenstein, L., *Philosophische Untersuchungen*, Suhrkamp, Frankfurt, 1984, 265.

potencialmente pública. Ora, qualquer linguagem, que seja potencialmente comunicável, não é privada no sentido que Wittgenstein pretende recusar. A recusa da linguagem privada é a recusa da possibilidade de uma linguagem sem critérios, o que talvez contribua para a compreensão do modalizador de necessidade em questão, com relação ao qual Bennett parece criar certo embaraço.

O segundo argumento estaria voltado para as condições segundo as quais podemos formular, publicamente, nossos estados internos. Aqui a proposta de Bennett se assemelha a de Strawson, na qual nossa linguagem pública para sensações adquire um caráter derivado. Vimos anteriormente que por trás desta interpretação estava a própria aceitação de uma linguagem privada, o que a torna incompatível com a proposta de Wittgenstein. No que concerne mais especificamente à interpretação de Bennett, acrescentaria ainda que Wittgenstein não parece dizer que a causa de nossas sensações é externa, mas apenas que o contexto e o comportamento, envolvidos nas manifestações de uma sensação, constituem os critérios pelos quais discriminamos experiências internas e aprendemos a aplicar-lhes um nome.

IV

Até aqui podemos concluir que tanto Strawson como Bennett discordam de Wittgenstein por considerarem possível exatamente aquilo que Wittgenstein pretende recusar, qual seja, a possibilidade de que sejamos capazes de identificar algo à revelia de critérios de identificação externos/objetivos. Do mesmo modo, confundem o seguir regras privadas com o agir de acordo com regras isoladamente. O que o argumento da linguagem privada pretende recusar é a possibilidade de regras privadas, ou seja, que não possam ser compreendidas por outros sujeitos, por conseguinte, que não sejam, nem mesmo potencialmente, públicas.

De acordo com esta interpretação podemos reconstruir o argumento da linguagem privada em duas etapas. A primeira, destinada a elucidar o modo como aprendemos a aplicar nomes para sensações, responderia a todos aqueles que recusam a possibilidade de uma linguagem pública acerca das sensações. Com esta etapa Strawson e Bennett parecem concordar. A segunda combateria o suposto elemento privado da linguagem das sensações, denunciando a ausência de critérios que tornassem possíveis definições privadas e afastando o recurso a processos mentais ocultos. Aqui encontramos, então, o passo decisivo do argumento, ou melhor, o argumento contra a possibilidade de uma linguagem privada propriamente dita.

Para elucidar esta segunda etapa, tomarei com ponto de partida a colocação do problema tal como ilustrada por Wittgenstein nas passagens referentes ao ‘diário das sensações’:

“(...) Quero escrever um diário sobre a repetição de uma certa sensação. Para tanto, associa-a com o símbolo ‘S’ e escrevo este signo num calendário, todos os dias em que tenho a sensação. - Observarei, primeiramente, que uma definição do

*signo é impronunciável. - Mas posso dá-la a mim mesmo como uma espécie de definição ostensiva! - Como? Posso apontar para uma sensação? - Não no sentido habitual. Mas falo ou escrevo o signo e ao fazê-lo concentro minha atenção na sensação; - aponto, pois, como que interiormente, para ela (...)"*²³

O que quer dizer tudo isso? Tem-se aqui em vista a suposição de que uma definição ostensiva limitada à apresentação de um signo juntamente com um objeto possa determinar o significado de um termo? Não apenas isto, mas supõe-se ainda que isto seja possível sendo o objeto algo interno, privado. Ora, para que pudéssemos correlacionar desta maneira um signo e uma sensação, seria necessário que ao menos a própria sensação já houvesse sido discriminada e neste caso, não seria a definição ostensiva a responsável pela discriminação, mas, ao contrário, ela própria já suporia tal discriminação. Quando o que se quer discriminar e a regra, ou seja, a sensação e o signo, são internos, privados, torna-se impossível qualquer critério de correção, pois nada distingue um e outro, não havendo, portanto, discriminação alguma.

Se toda definição é dada de acordo com critérios que permitam a correção do uso de um termo de forma a determinar seu significado, nos casos em que não dispomos de correção nenhuma definição pode ser fornecida, e o emprego de tal signo não obedecerá a qualquer regularidade, sendo, portanto indiferenciado, sem significado. Deste modo, o 'S' de nosso diário de sensações não possui significado algum, ou seja, com ele não somos capazes de nomear ou discriminar qualquer de nossas sensações, nem mesmo privadamente. Apenas acreditamos que 'algo' seja anotado, mas nossa simples crença tampouco produz qualquer distinção e não fornece, assim, as discriminações necessárias para que possamos diferenciar nossos próprios estados internos.

Usar um signo deste ou daquele modo significa usar discriminadamente, de acordo, portanto, com uma regra de uso. Quem aplica regras deve possuir critérios de correção, do contrário tratar-se-ia de uma regra inútil, pois tudo estaria de acordo com a mesma, ou seja, nada estaria sendo determinado. Em tais casos, não pode haver acerto nem erro, simplesmente porque não há sequer diferenciação. Quem diz utilizar um signo significativamente, deve admitir utilizá-lo de forma determinada, ou seja, discriminada, pois, do contrário, não estaria defendendo uma linguagem privada, mas supondo uma experiência indiferenciada, caótica, acerca da qual nada poderia ser dito.

Nosso interlocutor deverá conceder não apenas que uma linguagem pública seja possível, mas ainda que seja capaz de discriminar seus estados internos. Quem admite ser capaz de discriminar estados internos, deve admitir a posse de critérios. Numa tal situação, os critérios jamais poderiam ser igualmente internos, pois perderiam sua função, qual seja, produzir distinções. Se a possibilidade de

²³ Idem, ibidem, 258.

discriminarmos sensações está condicionada à posse de critérios externos, ou seja, objetivos, em que sentido tal linguagem ainda poderia ser dita relevantemente privada? O fato de que cada um possui sua própria sensação, e dispõe de um saber privilegiado acerca das mesmas²⁴, não as torna de modo algum incomunicáveis, nem tampouco dispensa a necessidade de critérios que nos permitam distinguir uma sensação dentre as demais.

Não basta, portanto, que suponhamos substituir sensações indeterminadas pela palavra ‘dor’, pois não podemos aplicar uma palavra senão quando podemos determinar seu uso, e isto significa que devemos ser, ao menos potencialmente, capazes de fornecer seus critérios de correção. A correção de tal uso só é possível se apelarmos a elementos independentes, caso contrário, não disporíamos do contraste necessário a qualquer diferenciação. A ausência de tais critérios em uma linguagem privada elimina a possibilidade de que seu defensor seja capaz de discriminar sensações ou estados internos, por conseguinte, resta-lhe tão somente uma experiência indiferenciada, em meio a qual nem mesmo uma suposta linguagem privada seria possível. Deste modo, ou bem estará condenado ao silêncio, ou terá que admitir que os critérios, nos quais se baseia sua possibilidade de discriminar estados internos, são independentes destes próprios estados, o que destitui o caráter privado de sua linguagem.

Conclusão:

Na primeira etapa desta exposição, procurei apresentar a origem pré-filosófica da crença em uma linguagem privada. A partir da crença em uma origem privada da linguagem, fiz emergir a tradicional questão cética: se o que confere sentido a nossas palavras são vivências, privadas, subjetivas, como poderemos garantir que ‘realmente’ possa haver uma linguagem comum? Se pudermos remeter todas as nossas experiências à vivência privada das mesmas, não haveria porque considerar problemática apenas a linguagem pela qual expressamos sensações. O sentido de toda e qualquer linguagem seria originalmente privado. Por conseguinte, já não haveria qualquer garantia de uma linguagem pública.

Para analisar o argumento contra a possibilidade de uma linguagem privada de Wittgenstein, adotei como ponto de partida a reconstrução do mesmo fornecida por Strawson e Bennett. Na interpretação de ambos, Wittgenstein parece incorrer em um equívoco. De acordo com primeiro, seu erro teria sido supor que não podemos identificar nossos próprios estados senão por recuso a critérios externos. Segundo Strawson uma sensação é identificada por quem a sente, e a necessidade de recorrermos a critérios só se coloca quando pretendemos tornar esta mesma sensação pública. Caberia, assim, apenas explicar a possibilidade de uma linguagem pública

²⁴ Ver E. Tugendhat, *Selbstbewusstsein and Selbstbestimmung*, Suhrkamp, Frankfurt 1979

acerca de sensações. Como solução o autor nos sugere que tomemos nossa própria experiência como modelo para atribuir sensações a outros sujeitos. Para Bennett, Wittgenstein estaria correto ao elucidar o modo pelo qual tornamos uma linguagem pública, mas se equivocaria ao supor impossível uma linguagem privada.

Contra Strawson e Bennett, defendi a tese de que o essencial do argumento da linguagem privada consiste em mostrar que os critérios pelos quais discriminamos nossas próprias sensações e estados psicológicos, os critérios pelos quais determinamos o significado de nossas expressões, devem poder ser, ao menos potencialmente, comunicáveis. Assim sendo, ainda que nossas vivências não possam ser compartilhadas, seu significado deve poder ser determinado mediante critérios objetivos, nem um pouco inefáveis, em nada transcendentais.²⁵

Partimos, assim, da capacidade de discriminar estados subjetivos ou sensações. A discriminação e a identificação de uma sensação só é possível se possuímos critérios. Critério é algo que permite estabelecer distinções, portanto, não pode confundir-se com aquilo a que deve distinguir. Resulta, assim, que nossas sensações ou experiências subjetivas em geral devem envolver critérios públicos de discriminação, o que elimina o suposto caráter privado de nossa linguagem acerca das mesmas.

Deste modo, resta ao defensor de uma linguagem privada apenas a possibilidade de retirar sua premissa, ou seja, recusar a capacidade de discriminar estados subjetivos. Porém, neste caso, suprimiria não só a necessidade de critérios, mas também a possibilidade de uma linguagem com sentido.

²⁵ Acerca da distinção entre a posse de uma sensação e o conhecimento da mesma, ver: P.M. Hacker, *Insight and Illusion*, Clarendon Press, Oxford 1986 e A. Kenny, *Wittgenstein*, Penguin Books 1973.

Causação Intencional

Um dos temas mais recorrentes da filosofia, em geral, e da filosofia da mente, em particular, é o da relação entre o mental e o físico. Retomada sob a forma da clássica distinção cartesiana entre *res cogitans* e *res extensa*, a discussão transita do plano ontológico para o epistêmico deixando atrás de si um leque de respostas mais ou menos incompatíveis e, sobretudo, um amplo espectro de discussões correlatas que tornam pungente sua solução. Para buscarmos uma introdução pré-filosófica ao tema, podemos reportar-nos às inúmeras formas cotidianas de descrição/interpretação do mundo no qual vivemos ou de nós mesmos. Explicamos a ocorrência de fenômenos físicos por recurso a eventos igualmente físicos, mas também por recursos a fenômenos de outro tipo. Para sermos mais generosos com o nosso senso comum, ou simplesmente para simplificarmos o já bastante amplo espectro deste debate, reportarei este segundo grupo de fenômenos aos fenômenos mentais. É neste sentido que vemos, por exemplo, a taquicardia ser atribuída, algumas vezes, ao aumento da adrenalina na corrente sanguínea e, outras, ao medo, desejo ou a outros “estados d’alma”. Da mesma forma, costumamos explicar fenômenos mentais por recurso a outros fenômenos mentais, mas também por recurso a fenômenos físicos. Dizemos que uma pessoa está apática ou triste por ter sofrido uma desilusão amorosa, mas também por estar sofrendo alterações hormonais. Dito assim de forma tão simplificada, tais explicações podem parecer disparatas e pouco verossímeis. O fato é que, verossímeis ou não, nosso dia-a-dia está repleto de tais explicações e não nos parece causar espanto que um mesmo fenômeno possa ser descrito e explicado de formas tão diversas. No âmbito epistêmico costumamos também acreditar que o amadurecimento da ciência nos trará, em cada caso, a explicação correta, desbancando as demais sob o título de credices. O que antes dizíamos ser possessão, passamos a interpretar como desequilíbrio psíquico/afetivo e caminhamos para a aceitação de que sejam casos de desequilíbrio orgânico. No mundo físico parecemos, hoje, depositar a chave explicativa inequívoca de todos os fenômenos.

No plano ontológico a discussão entre o mental e o físico leva-nos de imediato a duas questões: (1) São o físico e o mental entidades distintas? (2) Como entidades distintas podem manter relações causais? Dualistas, como Descartes, respondem afirmativamente à primeira questão, mas enfrentam dificuldades em responder a segunda. Assumindo que apenas entre entidades de um mesmo tipo podemos estabelecer relações causais, outros recusarão a perspectiva dualista, mas terão que explicar a aparente distinção entre o mental e o físico. As alternativas neste caso são diversas: alguns localizaram a distinção apenas no nível descritivo, ou seja, trata-se de um único tipo de fenômeno que pode ser descrito de formas distintas; outros defenderão uma relação de identidade, onde então os fenômenos mentais deixam de existir, ou de superveniência, onde os fenômenos mentais perdem sua eficácia causal.

Nesta exposição, pretendo apresentar a perspectiva de Searle como uma tentativa de resposta ao problema da relação causal entre o mental e o físico. Searle procura, dentro de uma proposta naturalista, resgatar a eficácia causal do mental. Neste sentido, ele recusa tanto o dualismo ontológico, como o reducionismo, a redução explicativa do mental ao físico, e o eliminativismo, ou seja, a completa eliminação do mental.

Para apresentar a proposta de Searle pretendo analisar o que o autor considera como sendo o característico e irreduzível do mental, a saber, a Intencionalidade²⁶. Situando então o problema como uma explicação e defesa da chamada causalidade Intencional, pretendo analisar as condições necessárias para que, segundo Searle, possamos atribuir satisfatoriamente eficácia causal a nossos estados Intencionais.

1. A Intencionalidade dos estados Intencionais:

A Intencionalidade é para Searle a propriedade dos estados ou eventos mentais pelos quais estes estão dirigidos-a, ou são acerca-de, objetos e estados de coisa no mundo. Um estado mental com tal característica será então dito estado Intencional. Cumpre lembrar que nem todos os estados mentais são estados Intencionais, posto que alguns estados mentais não possuem Intencionalidade, como por exemplo, alguns estados de ansiedade, exaltação e melancolia. Pela mesma razão, devemos também distinguir Intencionalidade e consciência, pois podemos ter consciência de nossa ansiedade sem que esta seja reconhecida como estando direcionada para algo, ou seja, como possuindo Intencionalidade. Da mesma forma, alguns estados Intencionais podem ser ditos inconscientes, no sentido em que posso não tê-los presentes no momento. Tenho muitas crenças nas quais não estou pensando atualmente, mas que certamente exprimem estados Intencionais que efetivamente tenho. Acredito, por exemplo, que minha mãe jamais esteve na China e sou capaz de expressar tal crença como resposta a uma indagação alheia, embora possa jamais ter refletido sobre a mesma antes.

Mas se ser um estado mental ou ser consciente não nos permite identificar algo como sendo um estado Intencional, de que critério podemos então nos servir? A resposta é imediata. Se a Intencionalidade se caracteriza pelo seu direcionamento a um objeto, um estado será Intencional todas as vezes que pudermos responder a perguntas do tipo: a que se refere? Sobre que é tal estado? Onde obtemos como resposta um objeto ou estado de coisas, algo, portanto, distinto do próprio estado Intencional em questão. De acordo com tal critério, são então estados Intencionais crenças, temores, esperanças, desejos, amor, ódio, etc.

²⁶ Estarei utilizando, como propõe Searle, letra maiúscula para a expressão Intencionalidade e expressões derivadas para evitar que a mesma seja confundida com um tipo específico de estado Intencional, a saber, as intenções.

Para melhor caracterizar nossos estados Intencionais e sua relação com os objetos ou estados de coisa para o qual estão direcionados, Searle propõe analisá-los a luz de sua interpretação dos chamados atos ilocucionários. Atos ilocucionários são atos de fala através dos quais expressamos nossas crenças, desejos, temores, dúvidas etc. São compostos de um conteúdo proposicional—aquilo em que se acredita, que se deseja, teme, dúvida etc.—e uma força ilocucionária correspondente ao modo psicológico expresso—crença, desejo, temor, dúvida etc. Em um ato de fala $F(p)$ dizemos que F é a força ilocucionária e p o conteúdo proposicional. Um mesmo conteúdo pode então vir acompanhado de forças ou modos psicológicos diversos. Tomemos, por exemplo, a expressão: “estar chovendo lá fora”. Com este mesmo conteúdo podemos expressar o *desejo* de que esteja chovendo lá fora, o *temor* de que esteja chovendo lá fora, a *crença* de que esteja chovendo lá fora etc.

Assim como todo o ato de fala possui um conteúdo proposicional e uma força ilocucionária, todo estado Intencional possui um conteúdo representativo e um modo psicológico. Estados Intencionais representam objetos ou estados de coisas no mesmo sentido em que os atos ilocucionários os representam, a saber, seu conteúdo se reporta a um estado de coisas no mundo. Na realização de cada ato ilocucionário com um conteúdo proposicional expressamos certo estado Intencional que é a condição de sinceridade do ato de fala. Em outras palavras, a realização do ato de fala é necessariamente uma expressão do estado Intencional correspondente. Se, pelo ato de fala da promessa, prometo trazer o livro na próxima semana, então por este mesmo ato expresso a intenção de trazer o livro na próxima semana. Se afirmar que hoje está chovendo, expresso a crença de que hoje esteja chovendo. Atos ilocucionários e estados Intencionais cujo conteúdo é uma proposição completa possuem também uma direção de ajuste. Descrições e asserções possuem, por exemplo, direção de ajuste mente-mundo, ou seja são satisfeitas quando a mente se ajusta ao mundo. Desejos e intenções possuem direção de ajuste mundo-mente, ou seja, são satisfeitas quando o mundo se ajusta à mente. Nos atos ilocucionários ou nos estados Intencionais o conteúdo determina suas condições de satisfação e o modo psicológico ou a força ilocucionária, a direção de ajuste.

Atos ilocucionários e estados Intencionais possuem condições de satisfação ou sucesso. Um ato de fala será satisfeito se e somente se (i) o estado psicológico expresso for satisfeito e (ii) as condições de satisfação do ato de fala e do estado psicológico expresso forem idênticas. As condições de satisfação de um estado Intencional são *internas* ao próprio estado. Não podemos conhecer um estado, como, por exemplo, acreditar em algo ou desejar algo, sem conhecer suas condições de satisfação. Saber o que é um determinado estado Intencional significa exatamente conhecer suas condições de satisfação, ou seja, conhecer (i) o que é representado por seu conteúdo proposicional e (ii) qual o modo psicológico do estado em questão. Conhecer qual é o objeto Intencional, ou o estado de coisas representado é, por sua vez, saber a que se refere, o estado Intencional. Se nada satisfaz a porção referencial

do conteúdo representativo, diremos que o estado Intencional é desprovido de objeto Intencional.

Tal explicação dos estados Intencionais e de seus objetos dispensa, portanto, o conhecimento de categorias ontológicas. A única condição externa ao próprio estado assumida por Searle é a de que o estado Intencional em questão esteja relacionado a outros estados Intencionais (o que ele chamará de “rede”) e se situe sobre o pano de fundo (*background*) de práticas e suposições pré-intencionais. Se não relacionarmos, por exemplo, a expressão de um estado Intencional de um ator com os demais estados Intencionais de seu personagem e o contexto narrativo em geral, do qual faz parte não apenas o roteiro da peça em questão, mas também seu autor, sua época, o público ao qual se dirige, os costumes vigentes etc., não seremos capazes de compreendê-lo. É neste sentido, portanto, que a rede e o *background* passam a integrar as condições para a compreensão de um estado Intencional.

Mas se as condições de satisfação de um estado Intencional são as condições impostas pela sua compreensão, em que medida aspectos do mundo físico, como por exemplo, as estruturas cerebrais, podem desempenhar algum papel na discussão acerca de tais estados mentais? Diante da dicotomia mente-corpo, Searle assume uma perspectiva não dualista, naturalista. Para ele, a solução não está em negar a existência de fenômenos mentais, mas em estimar adequadamente sua natureza biológica: os fenômenos mentais são ao mesmo tempo causados por operações do cérebro e realizados na estrutura cerebral.²⁷ A investigação de tais processos cabe, no entanto, às ciências empíricas. A base biológica dos fenômenos mentais não impede que possamos reconhecer características peculiares aos mesmos, tal como, por exemplo, a Intencionalidade, e que, com base em uma análise desta característica, possamos explicar satisfatoriamente a atribuição de significado a nossas práticas linguísticas e a nossas ações. A solução proposta por Searle será discutida na quarta parte desta exposição.

Vimos até aqui como podemos compreender um estado Intencional composto por um conteúdo proposicional e um modo psicológico. Os casos paradigmáticos são para Searle as crenças e os desejos. Ambos possuem como complemento uma proposição completa que representa certo estado de coisas. Compreender tais estados significa saber em que condições serão satisfeitos. O modo psicológico de cada um determina a sua direção de ajuste: mente-mundo, no caso das crenças, e mundo-mente, no caso dos desejos. O que devemos dizer, no entanto, no caso dos estados Intencionais que não possuem como conteúdo uma proposição e para os quais não podemos identificar diretamente uma direção de ajuste? Podemos reduzir outras formas de Intencionalidade a crenças e desejos? Aqui a hipótese adotada por Searle será a de que todos os estados Intencionais, mesmo aqueles sem direção de ajuste e

²⁷ Ver, J. Searle, *Intencionalidade*, Martins Fontes, São Paulo, 1995, X.

que não tem por conteúdo uma proposição completa contém crenças, desejos ou ambos e que, ao menos em diversos casos, a explicação da Intencionalidade será reportada a estes últimos.

Para chegarmos às crenças e desejos embutidos nos demais estados Intencionais devemos tomar um estado específico e nos perguntar em que devemos acreditar e/ou o que devemos desejar para estar no estado em questão. Se tivermos, por exemplo, esperança de que os conflitos no continente africano terminem, podemos então dizer (i) que acreditamos que haja conflitos no continente africano e que (ii) desejamos que tais conflitos cheguem ao fim. Se lamentarmos termos nos atrasado para o filme, então isto significa que (i) acreditamos ter chegado atrasados, mas que (ii) desejávamos chegar a tempo. Poderíamos prosseguir com o teste indefinidamente e traduzir nestes termos nossos temores, orgulho, vergonha, desapontamento etc. Ao invés disto, proponho que aceitemos a fórmula sugerida por Searle e passemos a outra questão: o que fazer nos casos em que a Intencionalidade, por não se reportar a estados Intencionais, não puder ser explicada através de crenças e desejos? Tal como os estados Intencionais, nossas experiências perceptivas e nossas ações se direcionam a objetos e estados de coisa no mundo, possuem, portanto Intencionalidade. Mas como podemos explicar seu modo de se relacionar com o mundo? Estamos agora diante de dois novos âmbitos de investigação: a Intencionalidade das experiências perceptivas e a Intencionalidade das ações.

2. A Intencionalidade das experiências perceptivas:

As experiências perceptivas são para Searle eventos mentais conscientes e se relacionam com suas condições de satisfação de modo diverso do das crenças e desejos. Compartilham com os estados Intencionais a propriedade da direcionalidade e possuem como características peculiares o imediatismo e a involuntariedade.

Inicialmente três semelhanças podem ser apontadas entre as experiências perceptivas, entre as quais elegeremos como modelo a experiência visual, e os estados Intencionais, exemplificado pelas crenças e desejos. Em primeiro lugar, em ambos os casos o conteúdo representacional equivale a uma proposição completa. A experiência visual é sempre de que determinado estado de coisas se verifica e não de um objeto. Ao relatar nossas experiências visuais dizemos, por exemplo, que “vemos que lá fora está chovendo” ou “que a árvore está florida” etc. O complemento é, portanto, uma proposição. A experiência visual, no entanto, não se limita a representar o estado de coisas percebido, mas, quando satisfeita, faculta-nos um acesso direto a este. Neste sentido, seu modo peculiar de representar um estado de coisas será chamado apresentação. Em segundo lugar, tal como as crenças e desejos, nossas experiências visuais tem direção de ajuste. Sua direção de ajuste se assemelha a das crenças, a saber: mente-mundo. Em terceiro, experiências visuais, assim como as crenças e desejos, são identificadas e descritas em termos do seu conteúdo

intencional. Nos exemplos acima teremos como conteúdo Intencional: “que lá fora está chovendo” e “que a árvore está florida”.

O conteúdo Intencional da experiência visual determina sob quais condições ela é ou não satisfeita. É parte de suas condições de satisfação que a experiência visual seja causada pelo restante de suas condições de satisfação, a saber, o estado de coisas ao qual se reporta. Em outras palavras, a compreensão do conteúdo Intencional da experiência visual requer: (i) que haja um estado de coisas no mundo e que (ii) este estado de coisas cause a experiência visual. Neste sentido, a experiência visual será dita autorreferente e apresentará também uma direção de causação, a saber: mundo-mente.

O aspecto sob o qual algo é percebido é determinado tanto pelo ponto de vista daquele que percebe como pelas demais características da situação perceptiva. Assim sendo, mais uma vez a rede e o *background* tornam-se determinantes do que é percebido. Podemos, por exemplo, expor duas ou mais pessoas aos mesmos estímulos visuais e obter com o experimento percepções visuais distintas. A diferença no resultado obtido poderá então ser explicada pelos estados Intencionais atuantes em de cada pessoa no momento da percepção e por fatores não Intencionais relacionados, por exemplo, a suas atividades, costumes, língua etc. Exercícios de figura e fundo amplamente utilizados na psicologia como “o coelho e o pato”, “a velha e a moça”, “o cálice e os dois perfis”, servem para ilustrar esta influência de outros fatores na determinação do conteúdo perceptivo. Em alguns casos, a influência de um estado Intencional pode chegar a se sobrepor à experiência visual, gerando diferentes condições de satisfação. Este é, por exemplo, o caso da nossa percepção do sol e da lua. Em ambos os casos, nossa experiência visual se confronta com a crença, amplamente reconhecida em nossa cultura, de que o sol ou a lua são infinitamente maiores que nós. Ao sermos indagados sobre o que estamos vendo, dificilmente cedemos a tentação de dizer que o objeto apresentado tem as dimensões de uma bola de futebol. As condições de verdade de uma experiência visual envolverão o perceptante e os componentes da percepção visual, a experiência visual e a cena, numa relação Intencional, causal. Rede, *background* e causação figuram, portanto, na determinação das condições de satisfação de uma experiência visual.

Ao caracterizar deste modo os fatores determinantes das condições de satisfação de nossas experiências perceptivas, Searle supõe responder a dois problemas clássicos da filosofia: (1) o problema da particularidade, ou seja, como objetos particulares podem fazer parte do conteúdo Intencional e (2) o problema do ceticismo, ou seja, como podemos garantir validade objetiva ao nosso conhecimento.

Vejam agora o primeiro caso. Como podemos identificar entidades particulares no conteúdo intencional da nossa experiência visual? Suponhamos que

estejamos diante de uma pessoa e que esta seja apresentada a nós como sendo Inês²⁸. Passam-se alguns dias e encontramos novamente alguém com as mesmas características físicas da pessoa a qual fomos apresentados como sendo Inês. Prontamente lhe estendemos a mão e cumprimentamos: “como vai, Inês?” A pessoa recua e diz termos cometido um engano. Surpresos, narramos o ocorrido a uma terceira pessoa e ela nos revela que Inês tem uma irmã gêmea. Nos dois momentos tivemos a experiência visual de uma pessoa. No segundo momento, no entanto, nossa identificação do objeto percebido fracassa. A que devemos atribuir tal fracasso, ou melhor, a que devemos atribuir o sucesso da experiência anterior? A resposta, então, será: ao reconhecimento de que o objeto em questão causa a experiência perceptiva e a correlação do objeto percebido com o conteúdo de outros estados Intencionais e/ou outras situações perceptivas. Para reconhecer uma pessoa como sendo Inês devemos ser, portanto, capazes de reportar o conteúdo da experiência perceptiva a outros conteúdos, tais como, por exemplo, a crença de que a pessoa que nos foi apresentada é amiga de Pedro, estudante de arquitetura e que esteve no mesmo local que nós uma semana antes. Em condições normais, ou seja, dada certa regularidade no mundo em que vivemos, quem satisfizer tais características será corretamente identificado como sendo Inês. As semelhanças físicas entre Inês e sua irmã gêmea não nos induzirão continuamente ao erro desde que possamos nos servir de outros conteúdos Intencionais que se refiram inequivocamente a Inês e não a sua irmã. A identificação de particulares torna-se assim possível graças ao reconhecimento da rede e do *background* como determinantes das condições de satisfação da experiência perceptiva.

Acabamos de ver como experiências visuais de tipo idêntico podem ter condições de satisfação distintas e como tais condições podem ser particulares e não gerais. Cabe-nos agora a pergunta inversa: como pessoas diferentes, com vivências diferentes podem ter estados Intencionais e experiências perceptivas com as mesmas condições de satisfação. De acordo com a perspectiva cética, nossa própria experiência não nos forneceria uma base suficiente para nos certificarmos de estarmos nos reportando a objetos e estados de coisas comuns. Segundo a argumentação cética: (i) para que eu esteja vendo um carro é necessário que minha experiência visual seja causada pelo carro. (ii) Devo saber que o carro causou a experiência visual e infiro da presença e da natureza de minhas experiências, como efeito, a existência e as características do carro, como causa. Para o cético, no entanto, uma inferência causal deste tipo jamais poderia ser justificada, uma vez que o único acesso que tenho ao carro é através da minha própria experiência. Por conseguinte, tomando como base a minha experiência visual, jamais poderei saber se o carro realmente causou minha experiência ou mesmo que haja algum carro.

²⁸ A construção deste exemplo se baseia no exemplo de Sally fornecido por Searle no capítulo 2 de *Intencionalidade*, Martins Fontes, São Paulo, 1995.

Para Searle, tal argumento fracassa ao supor, já na formulação do problema, que nosso conhecimento da relação causal seja inferencial: “Não infiro (afirmo ou descubro) que o carro está causando a minha experiência visual. Simplesmente vejo o carro.(...) O conhecimento de que o carro causou minha experiência visual deriva do conhecimento de que o vejo, e não o contrário”.²⁹ Não fazemos uma inferência causal da experiência visual para a existência do carro. A experiência visual não representa a relação causal como algo que exista independente da experiência. Não estamos lidando com duas coisas uma das quais é a evidência da outra, mas percebemos uma única coisa e, ao percebê-la temos uma experiência perceptiva. Ter uma experiência visual é estar em condições de reconhecer direta ou imediatamente a relação causal entre o estado de coisas percebido e a experiência visual do mesmo.

É possível que a ocorrência da experiência não nos forneça a garantia da existência do objeto, mas isto não nos induz a pensar que a relação entre nossa experiência e o objeto seja inferencial. Podemos nos equivocar com relação ao objeto da nossa experiência e reportar nosso engano à ausência de *background* ou a interferência de crenças, desejos ou outros estados Intencionais. Adotando um ponto de vista filosófico, podemos ainda estar em busca “do que as coisas são independentemente de serem percebidas”, ou seja, da “coisa em si”. Neste caso, de fato não teremos mais nada a provar. Apenas repetiremos, com Kant, que ou bem o que as coisas são deve ser entendido como relativo à nossa capacidade de tomá-las como nosso objeto de conhecimento—nos termos de Searle, nossa capacidade de receber os inputs causais de um mundo—ou se tornam, para nossa investigação, irrelevantes, pois já não seriam aquilo a que nos direcionamos, ou seja, nosso objeto Intencional.

3. A Intencionalidade das ações:

Tal como muitos de nossos estados mentais e como nossas experiências perceptivas em geral, a Intencionalidade pode ser também atribuída às ações. Toda ação pode ser descrita como composta por um componente Intencional, a saber, a experiência de agir, e um evento, uma transformação no mundo. Tal transformação ou evento é aquilo que é visado pela intenção que leva a cabo a ação. A esta intenção chamaremos intenção em ação. Toda ação possui uma intenção em ação, sem a qual a realização de um evento qualquer não poderia ser reportada a um estado Intencional do agente. Compreender ou ter consciência de uma ação significa, portanto, compreender a relação entre a intenção em ação e o evento realizado. Tal como descreve Searle:

Minha ação de erguer o braço tem dois componentes, a intenção em ação e o movimento do braço. Tire-se o primeiro, não se tem uma ação, mas apenas o movimento; tire-se o segundo e não se tem êxito, mas apenas um esforço frustrado.

²⁹ J. Searle, *Intencionalidade*, Martins Fontes, São Paulo, 1995, 103.

*Não há ações, nem mesmo não-intencionais, sem intenção, pois toda ação tem uma intenção em ação como um de seus componentes.*³⁰

Para conhecermos a intenção em ação basta que sejamos capazes de responder à questão: o que você está fazendo agora? Isto significa: conhecer as condições de satisfação da ação. O conteúdo Intencional da intenção em ação e o conteúdo Intencional da experiência de agir são idênticos. A experiência de agir, ou seja, o componente Intencional da ação é simplesmente o que é expresso pela intenção em ação.

A experiência de agir possui, tal como as experiências perceptivas, uma direção de ajuste e uma direção de causação. Sua direção de ajuste será: mundamente e sua direção de causação: da experiência de agir para o evento. O método proposto para a identificação do conteúdo Intencional é interrogar o que deve ocorrer para que o conteúdo Intencional seja satisfeito. Identifica-se, assim, a Intencionalidade por suas condições de satisfação. Quanto à relação de causação exemplificada nas ações, diremos que a intenção em ação causa o evento. Neste sentido, estamos mais uma vez diante de um caso de autorreferencialidade causal. Ser consciente de uma ação significa (i) saber reportar o evento do qual a ação é composta ao componente Intencional da mesma, ou seja, à intenção em ação e (ii) reconhecer na intenção em ação a causa do evento em questão.

Searle faz questão de distinguir aqui a intenção em ação da chamada intenção prévia. A intenção prévia é aquela formada antes da ação, envolve um plano de ação, cálculos, estratégias etc. É claro que grande parte de nossas ações são levadas a cabo por uma intenção prévia que causa nossa intenção em ação. O importante é que possamos reconhecer que nem todas as nossas ações são acompanhadas de intenções prévias, mas mesmo assim são ações das quais somos conscientes e pelas quais podemos responder. Ações como escovar os dentes ao levantar, dirigir-se ao ponto de ônibus na hora de ir para o trabalho, lavar as mãos ao chegar da rua, beijar nossos filhos ao entrar em casa etc., são ações sobre as quais raramente refletimos previamente, contudo, se formos interrompidos em qualquer uma delas e indagados sobre o que estamos fazendo, saberemos sem dificuldade revelar o componente Intencional de nossa ação, a saber: a intenção em ação.

A intenção em ação é para Searle o que distingue ações de meros atos. Na ocorrência de atos não identificamos uma intenção em ação como causa do evento realizado. Automatismos são os exemplos mais comuns de atos. Nestes casos, ao sermos indagados sobre o que estamos fazendo, costumamos responder que “nada”, ou seja, não reconhecemos em nós, em nossos estados Intencionais, o fator causal relevante para a explicação dos movimentos em questão.

³⁰ J. Searle, *Intencionalidade*, Martins Fontes, São Paulo, 1995, 148.

A descrição de o nosso agir pode envolver ainda (i) intenções complexas, ou seja, pode nos reportar a uma rede de intenções prévias atuando como causadores³¹ de nossa intenção em ação e (ii) situações compostas por ações intencionais e não intencionais. Este último caso é bem ilustrado pelo drama de Édipo: Édipo casa-se com Jocasta. Casar-se com Jocasta é uma ação Intencional. Édipo sabia o que estava fazendo, ou seja, teria sido capaz de identificar a intenção em ação presente no seu agir. Tragicamente, casar-se com Jocasta significa também: casar-se com sua mãe. Tal ação pode, contudo, ser reconhecida como não-intencional. Édipo não reconheceria a intenção de casar-se com sua própria mãe como a intenção em ação que determinou o seu agir. Tudo o que podemos dizer é que a ação completa continha elementos que faziam parte da intenção em ação e outros que não faziam. Isto significa que alguns elementos faziam parte das condições de satisfação da intenção em ação e, por conseguinte, da ação propriamente dita e outros não. Em muitos casos, podemos até mesmo ter consciência de algumas consequências de uma ação, sem que as mesmas sejam incluídas nas condições de satisfação da intenção. Quando tomamos, por exemplo, a decisão de realizar um tratamento dentário, sabemos de antemão que horas não muito agradáveis nos aguardam. Nos casos comuns, a intenção que nos mobiliza, no entanto, não invoca tais sensações desprazerosas, mas sim o tratamento.

A correlação entre elementos Intencionais e elementos não-Intencionais envolvidos na compreensão de nossas ações é bastante explorada nas tragédias exatamente por despertar em nós esta sensação de estarmos diante de algo ao mesmo tempo estranho e familiar, de sermos a um só tempo agentes e vítimas dos acontecimentos. O que efetivamente distingue a situação de Édipo e o nosso dia-a-dia? A resposta será mais uma vez o recurso às regularidades planejáveis que constituem de certo modo o *background* de nossas ações ordinárias. Para ressaltar este aspecto, Searle dispõe-se a analisar casos em que a relação entre ações e intenções nos conduz a certos paradoxos.

O primeiro caso é sugerido por Chisholm³²: Bill deseja matar seu tio, do qual é herdeiro, a fim de enriquecer. Enquanto dirige, Bill planeja o assassinato e este pensamento o deixa tão nervoso que acaba provocando um acidente. Bill atropela e mata um pedestre, que vem a descobrir ser o seu próprio tio. Bill tinha a intenção prévia de matar o tio. Porém, não podemos dizer que esta intenção prévia tenha determinado a intenção em ação que levou à morte de seu tio. A intenção em ação de

³¹ Searle propõe uma distinção entre causadores e outros elementos de uma relação causal. Os causadores são os aspectos causalmente relevantes que aparecem em uma relação causal. São aqueles aos quais devemos reportar o nexos causal. Nem todos os relatos de relações causais são contudo relatos envolvendo causadores, ou seja, são relatos que identificam os elementos causalmente relevantes.

³² R.M., Chisholm, "Freedom and Action", in K. Lehrer (ed.) *Freedom and Determinism* (New York: Random House, 1966), p. 37.

Bill, no momento do acidente, não era matar seu tio, mas simplesmente “estar dirigindo”. Neste caso há uma quebra na cadeia causal que faz com que não possamos atribuir à ação presente de Bill a intenção de matar seu tio, ou seja, o poder causal sobre o evento ocorrido.

O segundo caso, semelhante ao anterior, é extraído de Davidson³³: Um alpinista deseja livrar-se do fardo e do perigo que representa para ele ter que segurar o companheiro. Este pensamento o deixa tão atormentado que ele solta a corda que o prendia ao outro alpinista, deixando-o cair. Aqui, mais uma vez, podemos dizer que não foi a intenção prévia de se livrar de um fardo que fez com que o alpinista deixasse cair seu companheiro. Havia uma intenção prévia e ocorreu um evento que coincide com a mesma. A intenção em ação no momento do acidente não estava, contudo, relacionada causalmente à intenção prévia e a queda do companheiro, neste caso, está contabilizada entre os elementos não intencionais da ação. O alpinista queria soltar o companheiro, assim como Bill queria matar seu tio. Este desejo entrou em relação com outros fatores que levaram a realização de uma ação composta por elementos intencionais e elementos não intencionais. Há, no entanto, uma quebra nonexo causal que faz com que movimento realizado não possa ser reportado ao desejo de ambos. Para nos certificarmos disto basta indagarmos pela intenção em ação do momento. Bill não nos responderá que está matando o seu tio, assim como o alpinista não nos dirá que está soltando seu companheiro, embora, nos dois casos isto esteja efetivamente acontecendo, porém enquanto parte dos elementos não intencionais de suas ações.

O terceiro caso, distinto dos anteriores, é fornecido por Bennett³⁴: Um homem tenta atirar em uma determinada pessoa para matá-la. Ele erra o alvo, mas o estrondo do tiro dissipa uma manada de porcos selvagens que acaba matando, pisoteada, a pessoa em questão. Neste caso, a intenção prévia causa a intenção em ação, mas esta última não consegue realizar o movimento desejado, ou seja, a ação não é satisfeita. A vítima é morta, mas não podemos reportar sua morte à intenção em ação. A quebra aqui se dá entre a intenção em ação e o evento. A intenção em ação seria satisfeita se o matador tivesse acertado a vítima com um tiro. Seu tiro causou outros eventos que levaram por fim à morte da vítima. A quebra de conexão entre as intenções e os eventos torna o resultado ao nível do agente tão fortuito como teria sido caso um raio tivesse atingido a vítima no momento em que o tiro foi disparado. O efeito do tiro sobre a manada é apenas mais um elemento não intencional da ação realizada.

4. Causação Intencional: o cérebro e a mente

³³ D. Davidson, “Freedom to act”, in T. Honderich (ed), *Essays on Freedom of Action* (Londres, Henley e Boston: Routledge & Kegan Paul, 1973), pp. 153-4.

³⁴ Citado por D. Davidson in T. Honderich (ed), *op.cit.*, pp. 152-3.

Os três exemplos acima mencionados envolvem (i) ou bem uma falha do conteúdo Intencional em atuar causalmente, tal como no último caso, (ii) ou uma falha de regularidade planejável nas relações causais do estado Intencional, tal como nos dois primeiros exemplos. Assim sendo, tais casos nos permitem discriminar ao menos duas condições necessárias para que possamos reconhecer relações causais entre eventos e estados mentais. A primeira condição é a de que o conteúdo Intencional atue eficientemente sobre os aspectos Intencionais da ação. A segunda, exige, como parte do *background*, que haja ao menos certo grau de regularidade planejável. O conteúdo Intencional deve ser um aspecto causalmente relevante e deve exemplificar uma regularidade planejável.

Com isso, Searle pretende fornecer uma explicação da causalidade que, ao contrário da tradição, nos permita reconhecer diretamente o nexos causal. Ou seja, tanto na percepção como no agir, a relação de causalidade é experienciada diretamente e reconhecê-la está entre as condições de satisfação de ambas. No caso das ações, a causalidade é parte do conteúdo da intenção em ação. Se a intenção em ação não causa o resto das condições de satisfação, a intenção não é satisfeita. No caso da percepção, a causalidade é parte das condições de satisfação do conteúdo perceptivo. Se não reconhecemos o objeto ou estado de coisas percebido como causa da nossa experiência perceptiva, não nos reconhecemos como tendo uma experiência perceptiva. É, portanto, parte do conteúdo da nossa experiência ser causa de um evento no mundo ou ser causada por um estado de coisas no mundo. Ao percebermos o mundo ou ao agirmos sobre ele temos estados Intencionais autorreferentes. A relação causal é interna, ou seja, é parte do conteúdo—e não do objeto—destas experiências.

Partindo desta descrição de uma relação causal, podemos agora, retomar a pergunta feita na primeira parte desta exposição: como podemos atribuir poderes causais a entidades não-Intencionais? Vimos, naquela etapa, que a atribuição de poderes causais ao mundo físico não é para Searle problemática. Ao contrário, o autor insiste em uma naturalização dos estados mentais, segundo a qual estes seriam não apenas realizados no cérebro, mas também causados por operações cerebrais. A experiência visual, por exemplo, é, segundo Searle, causada pelo funcionamento do cérebro em resposta a estímulos ópticos externos e também realizados na estrutura do cérebro. Dois fenômenos podem, assim, estar relacionados tanto pela causalidade, como pela realização. O importante é que isto aconteça em níveis distintos de descrição.³⁵ Para ilustrar estes dois tipos de relação, Searle elege como exemplo a água: a relação entre o comportamento molecular e as características físicas da água é um exemplo de causalidade. Ao descrevermos esta mesma substância como líquida, não estamos acrescentando-lhe uma propriedade adicional, mas apenas fornecendo outra descrição

³⁵ Ver, J. Searle, *Intencionalidade*, Martins Fontes, São Paulo, 1995, 369.

de uma propriedade que se realiza na estrutura molecular da substância em questão.³⁶ Searle admite, ainda, que em cada nível de descrição podemos também reconhecer relações causais distintas. Em suas próprias palavras:

*A explosão é causada pela descarga de uma faísca da vela, mesmo sendo a descarga e a explosão causadas por fenômenos em um micronível e nele realizados, nível de descrição este em que termos como 'descarga' e 'explosão' são inteiramente inadequados. Do mesmo modo, quero dizer que a intenção em ação causa o movimento corporal, ainda que a intenção em ação e o movimento corporal sejam causados em uma microestrutura e nela se realizem, estrutura em cujo nível termos como 'intenção em ação' e 'movimento corporal' são inadequados.*³⁷

Podemos afirmar, portanto, que toda a explicação fornecida até aqui do modo como a Intencionalidade pode tomar parte em relações causais não exclui a possibilidade de fornecermos para os mesmos fenômenos uma explicação igualmente válida que não contenha elementos Intencionais. Tais explicações não são excludentes, já que em cada caso estaremos enfocando aspectos irreduzíveis uns aos outros. Tais considerações são, por assim dizer, não a conclusão da proposta de Searle, mas hipóteses de trabalho não-questionadas ou premissas assumidas pelo autor. Contra aqueles que recusassem esta dupla forma de descrição e defendessem uma eliminação da explicação mentalista, Searle só poderia responder reiterando a peculiaridade de aspectos mentais como, no caso, da Intencionalidade, e mostrando a irreduzibilidade de uma explicação Intencional a explicações segundo leis meramente físicas. Contra os que recusam a possibilidade de explicarmos de acordo com leis físicas fenômenos relacionados a nossa vida mental, Searle pode apenas reiterar a promessa naturalista e deixar que a pesquisa científica revele relações que hoje ainda não podemos entrever.

Para concluir esta exposição gostaria apenas de indicar um argumento contra a tese da dupla explicação. Com isso, não pretendo negar, de antemão, nem a peculiaridade dos aspectos mentais, nem seus poderes causais, mas apenas apontar para as dificuldades reais de uma proposta conciliatória. Trata-se do princípio da exclusão explanatória de Kim.³⁸

Suponhamos, como propõe Searle, que duas explicações completas sejam fornecidas para um mesmo evento *e*: (i) uma delas invocando para *e* uma causa/explicação mental *m* e (ii) outra invocando uma causa/explicação física *p*. Para que ambas sejam verdadeiras é necessário que *e* seja causado por *m* e por *p*. Sendo ambas as explicações, por hipótese, completas, tanto *m* quanto *p* devem ser causas

³⁶ Ver, J. Searle, *Intencionalidade*, Martins Fontes, São Paulo, 1995, 368.

³⁷ J. Searle, *Intencionalidade*, Martins Fontes, São Paulo, 1995, 373.

³⁸ Para uma reconstrução e defesa do Princípio da exclusão explicativa, ver: Fuhrmann, A. e W. Mendonça, "Causas Excludentes", Manuscrito não-publicado, 2001.

completas de *e*. Segue-se, de acordo com Kim, que *m* e *p* não podem ser independentes uma da outra. Se efeitos físicos puderem, então, ser reportados a causas físicas completas, causas mentais serão inoperantes ou eficientes apenas em um contexto de sobre determinação. Neste último caso, no entanto, já não seriam elas mesmas causas completas ou independentes.

Se o argumento de Kim estiver correto, a possibilidade de fornecermos uma explicação causal física completa para um determinado fenômeno, possibilidade esta na qual Searle aposta, colocaria ao menos sob suspeita uma explicação mental, não-fisicalista do mesmo fenômeno. Se aceitarmos a possibilidade de uma dupla descrição, por conseguinte, uma dupla explicação, teremos ainda que investigar a relação e/ou hierarquia entre os dois tipos de explicação em questão. As alternativas seriam mais uma vez (i) ou bem o reconhecimento de uma relação de identidade, onde então a peculiaridade do mental seria eliminada, (ii) ou o reconhecimento de uma relação de superveniência, onde ao menos a eficácia causal do mental seria suprimida. Estaríamos, assim, diante das duas perspectivas que Searle procurou evitar: o eliminativismo, ou seja, a completa eliminação do mental ou a redução explicativa do mental ao físico.

O Fisicalismo e o Mistério da Consciência Fenomenal: observações sobre a teoria PANIC de Michael Tye*

Em seu livro *Ten Problems of Consciousness* Michael Tye pretende compatibilizar a perspectiva fisicalista com o reconhecimento de certas propriedades características dos chamados objetos fenomenais. Entre tais propriedades, as que representam maiores dificuldades para o fisicalista são: privacidade, necessário pertencimento a um sujeito e perspectividade. Tye propõe uma teoria original que caracteriza o caráter fenomenal da experiência consciente essencialmente como conteúdo intencional (representacional) não-conceitual. Tal conteúdo deve estar pronto ou em condições de ser utilizado (*poised*) pelo sistema cognitivo conceitual que gera atitudes proposicionais. A consciência fenomenal seria assim o *input* intencional de processos que têm como *output* outros estados intencionais de conteúdo proposicional (*beliefs* e *desires*), mas seria ela mesma não-conceitual. Nas palavras de Tye: “I call this the PANIC theory of phenomenal character: phenomenal character is one and the same as Poised Abstract Nonconceptual Intentional Content” (137). Com base nesta teoria, Tye se propõe ainda fornecer uma resposta adequada ao desafio apresentado por um dos mais importantes argumentos contra o fisicalismo ou, mais especificamente, contra a possível redução das propriedades fenomenais a propriedades físicas, a saber, o chamado argumento do conhecimento ou o “caso de Mary” formulado originalmente por Frank Jackson.

Nesta exposição, pretendo reconstruir (1) a dificuldade apresentada para o fisicalismo pelos chamados objetos fenomenais; (2) as soluções discutidas por Tye e, finalmente, (3) o argumento do conhecimento à luz da teoria PANIC do caráter fenomenal.

Por fisicalismo podemos compreender, em linhas gerais, a perspectiva segundo a qual tudo o que existe ou que pode ser conhecido faz parte do mundo físico. Físico, por sua vez, poderá ser compreendido como sendo (i) tudo o que pertence ao domínio da física, química, biologia molecular e neurofisiologia; (ii) tudo o que pode ser descrito por uma teoria adequada à explicação de seres não-vivos; (iii) tudo o que é constituído ou realizado por algo no domínio microfísico.

Todavia, qualquer que seja a definição preferida de “mundo físico,” há um conjunto de propriedades *sui generis* aplicáveis de fato a um grupo específico de objetos, mas que, pelo menos à primeira vista, *não podem* ser atribuídas significativamente aos objetos físicos. A impossibilidade neste caso parece ser “semântica:” não faria sentido atribuir as propriedades *sui generis* aos objetos físicos. É neste sentido que pareço ser o sujeito único da minha dor e que a mesma pareça só

* Agradeço a Wilson Mendonça, sem cuja contribuição este trabalho não teria sido realizado.

existir quando referida a mim. Dizemos, por exemplo, que “ninguém mais pode sentir a minha dor,” assim como também acreditamos que dor, tristeza, alegria e demais estados de espírito só existem enquanto são vivenciados ou enquanto pertencem a um sujeito. Minhas experiências perceptivas conscientes, minhas sensações corporais, emoções, paixões e humores parecem reportar-se a mim de tal forma que meu carro e minha roupa, por mais personalizados e relacionados à minha personalidade que eu queira interpretá-los, efetivamente não o fazem. Outra pessoa poderia ter o meu carro e a minha roupa, que, de resto, não deixam de existir tão logo deixem de ser posse de alguém. A “monopolização” de fenômenos mentais experienciais por um sujeito, por um “eu,” parece assim caracterizar uma forma de pertencimento intrínseco alheia aos objetos que estão contingentemente sob meu poder. Não deveríamos, então, distinguir os objetos aos quais tenho um acesso privilegiado, enquanto sou aquele que os *vivencia intrinsecamente*—isto é, os objetos que só podem ser “tidos” e conhecidos na *minha* perspectiva experiencial e que não subsistem a não ser em relação com sujeitos de experiência—de outros objetos aos quais me relaciono simplesmente porque sou aquele que os *tem* contingentemente? Se as propriedades características deste grupo de objetos, isto é, as propriedades da privacidade, do necessário pertencimento a um sujeito e da perspectividade não puderem ser atribuídas a objetos físicos, não deveremos então admitir a existência de entidades não-físicas?

É possível mostrar que isto implica reabilitar o dualismo cartesiano mente-corpo ou, pelo menos, adotar uma forma de epifenomenalismo a propósito dos objetos fenomenais. Sob a suposição plausível e aceita por todos que o mundo físico é causalmente fechado—as causas completas de eventos físicos incluem somente outros eventos físicos—isto nos deixaria impossibilitados de estabelecer relações causais entre o grupo de eventos aos quais supomos pertencer nossas sensações, emoções, paixões e humores, por um lado, e os eventos físicos, por outro. Na melhor das hipóteses, os eventos fenomenais poderiam ser efeitos, mas jamais causas de eventos físicos. Mas estaremos realmente dispostos a admitir que estados fenomenais sejam causalmente irrelevantes?

Para evitar tal conclusão, Tye analisa três alternativas:

1. Mostrar que os objetos fenomenais *não* possuem realmente as propriedades *sui generis* mencionadas;
2. Mostrar que os mesmos não existem e, por conseguinte, que tais propriedades não se aplicam;
3. Reconhecer a existência de objetos fenomenais e de suas propriedades características, negando, porém, que estas não se apliquem também aos objetos físicos.

A primeira alternativa é defendida com base em “experimentos” que apontam para a possibilidade de um mesmo objeto fenomenal ser possuído por diferentes sujeitos: o caso de Siameses; o do “eu replicante” (Zuboff) e do “cérebro clivado” (Unger). Os experimentos em questão são experimentos de pensamento nos quais algumas situações extremas são descritas, ao mesmo tempo em que se faz um apelo às intuições que guiam a aplicação dos nossos conceitos de experiência. A dificuldade aqui consiste basicamente no fato de que não podemos provar assim a impossibilidade de contraexemplos. Além disso, tais experimentos questionariam somente o suposto pertencimento intrínseco dos objetos fenomenais, deixando intactas as demais propriedades *sui generis* que geram problemas para o fisicalista.

Analisemos, agora, a segunda alternativa, ou seja, a recusa dos objetos fenomenais. Aqui é defendida a tese de que não há propriamente objetos fenomenais aos quais estaríamos relacionados na nossa experiência. O que existe é sempre uma experiência no sentido de um processo em nós, descrito, de forma enganadora, *como se* estivéssemos confrontados, neste processo, com objetos especiais. Não há, por exemplo, experiência *da* dor se isto for interpretado como a existência de um objeto chamado “dor” com o qual eu me confrontaria na minha experiência, mas sim uma experiência *dolorosa* detectada por alguns sensores internos e expressada verbalmente, pelo menos às vezes, mediante o uso de um verbo transitivo direto. Outras expressões possíveis e também em uso sugerem uma ontologia mais austera. Considere, por exemplo, a expressão primitiva que consiste no grito espontâneo de dor. Que expressão deve ser considerada a mais fiel aos fatos independentes? A expressão verbal da minha experiência “eu tenho dor” sugere uma relação real (*ter*) e dois objetos relacionados: *eu* e *minha dor*. Mas nós conhecemos vários outros casos nos quais a gramática superficial induz ao erro. A tese eliminativista a propósito dos objetos fenomenais diz, portanto, que no mundo não existem sensações que poderiam ser *tidas* por nós. Existem somente organismos que passam por certos processos (experiências) e que exprimem tais processos de forma enganadora quando dizem “eu tenho dor.” A expressão verbal perfeitamente adequada “está doendo,” por exemplo, não sugere objetos especiais e simplesmente qualifica adverbialmente um processo. O grau de plausibilidade da posição eliminativista aumenta quando consideramos processos não-experienciais. A descrição inocente “Pedro tem uma voz estridente” não nos levaria a afirmar enfaticamente a existência de dois objetos ligados pela relação *ter*: Pedro e sua voz. O que há verdadeiramente é que Pedro é o sujeito de um processo, de um evento com a característica da estridência. A nominalização “a voz estridente de Pedro” não se refere a um objeto especial ao lado de Pedro. Por que as coisas deveriam ser diferentes no caso da “dor intensa de Pedro”?

Evidentemente, o eliminativista a propósito dos objetos fenomenais ainda tem de dar uma resposta satisfatória à questão: a que devemos referir, em última instância, nosso discurso acerca das sensações? Se não há um objeto fenomenal, o que torna, então, verdadeiro nosso discurso fenomenal? A principal tentativa de resposta é fornecida pela chamada teoria adverbial das sensações ou experiências, na qual

podemos distinguir duas vertentes: (A) a teoria do operador de predicados (*the predicate operator theory*) e (B) a teoria do predicado de eventos (*the event predicate theory*). De acordo com a primeira, o advérbio modifica um predicado mais simples e funciona como um operador que introduz uma *nova* propriedade. Por exemplo, “Maria fala alto” pode ser analisado como “Maria tem a propriedade de falar alto.” Neste caso a função adverbial realizada pela palavra “alto” opera no predicado “falar” e introduz uma nova propriedade “falar alto,” *distinta* da propriedade mais simples “falar” e atribuída como um todo a Maria. De acordo com a segunda vertente da teoria adverbial, o advérbio deve ser compreendido como um adjetivo que se aplica a um evento. Por exemplo, “Maria fala alto” quer dizer que há um evento do qual Maria é o sujeito e tal evento tem a propriedade de ser alto.

Uma dificuldade para a primeira vertente da teoria adverbial aparece quando se tenta esclarecer enunciados que mencionam múltiplos objetos fenomenais. Intuitivamente, os enunciados “eu vejo um quadrado azul e um círculo vermelho” e “eu vejo um quadrado vermelho e um círculo azul” são distintos, mas qualquer tentativa de transcrever adverbialmente a referência aos objetos fenomenais em questão eliminaria a diferença—além de ser extremamente artificial em Inglês e mais ainda em Português: “eu estou vendo azuladamente e quadradamente e avermelhadamente e circularmente”.

A segunda vertente da teoria adverbial também tem seus problemas. Ao tratar advérbios como (adjetivos) qualificativos de eventos, ela é levada a identificar, em geral, o número de eventos experienciais com o número de objetos fenomenais mencionados na expressão verbal da experiência. Isto implica em particular: múltiplos objetos fenomenais *na mesma modalidade* (por exemplo, visão), múltiplos eventos experienciais. Quando as modalidades são diferentes, *não é* contraintuitivo falar de múltiplas experiências simultâneas. Eu posso ser simultaneamente o sujeito da experiência de ouvir a música e da experiência distinta de ver as luzes. Mas quando eu vejo um quadrado vermelho e um círculo azul ou tenho simultaneamente duas *afterimages*, o que ocorre—de um ponto de vista intuitivo—é que eu tenho *uma* experiência com múltiplos objetos fenomenais. Outro problema para a teoria do predicado de eventos aparece no fato de que ela transfere para os eventos experienciais propriedades dos objetos fenomenais, o que resulta, por sua vez, em contextos linguísticos estranhos. Assim, a referência à *afterimage* azul é eliminada e em seu lugar surge um “evento (experiencial) azul.” Não está claro, porém, o que possa significar dizer que um evento é azul.

Tye considera que o resultado líquido da apreciação crítica da teoria eliminativista apoiada pela teoria adverbial da experiência é que ela simplesmente desloca, sem vantagens explanatórias, o problema das propriedades características do âmbito dos objetos fenomenais para o âmbito da experiência concebida como evento. Para evitar a conclusão epifenomenalista, resta, assim, a terceira alternativa: reconhecer a existência de objetos fenomenais e de suas propriedades, negando que

estas não se apliquem aos objetos físicos. Tal alternativa será, finalmente, a adotada por Tye. Ela consiste em identificar os objetos fenomenais com eventos particulares ou *state tokens* físicos, aos quais as propriedades de privacidade, pertencimento a um sujeito, e perspectividade podem ser aplicadas. É verdade que tais propriedades não podem ser atribuídas a objetos físicos no sentido de *coisas* físicas localizadas espaço-temporalmente. Mas Tye procura mostrar que há eventos físicos não-controversos que instanciam também pelo menos algumas das propriedades problemáticas. Elas deixariam assim de ser *sui generis* e nós poderíamos reintroduzir os objetos fenomenais, concebidos, agora, como eventos, na teoria fisicalista da consciência fenomenal.

Como devemos, porém, conceber os eventos? Tye considera inicialmente a definição de eventos sugerida por Goldman e Kim, a saber: entidades complexas, analisáveis em objetos e propriedades temporalmente exemplificadas. De acordo com esta concepção, eventos são *constituídos* por objetos e propriedades instanciadas aqui e agora pelos objetos, pelos “sujeitos” destes eventos. O evento designado pela expressão “a erupção do Vesúvio no ano de ...” envolve o Vesúvio (o sujeito do evento) e uma propriedade complexa instanciada num certo lugar (no monte Vesúvio, no sul da Itália) e numa certa época (no ano ...). É importante notar que esta teoria tem a vantagem de resolver imediatamente o problema do pertencimento necessário, que parecia justificar o isolamento dos objetos fenomenais num reino próprio não-físico. A erupção do Vesúvio é, necessária- truísticamente, *do* Vesúvio, assim como a dor de Maria é *de* Maria. Outro vulcão não pode ser o “sujeito” da erupção do Vesúvio no ano E, no entanto, a erupção em questão é uma coisa física. Isto mostraria que a propriedade do pertencimento necessário não é a linha que divide o físico do não-físico. O mesmo vale para as características normalmente associadas a esta propriedade. Se os objetos fenomenais não podem sobreviver à aniquilação dos sujeitos experienciais que os têm, erupções vulcânicas tampouco existem sem vulcões.

Aos olhos de Tye, esta solução elegante para o problema do pertencimento necessário é, sem dúvida, um ponto positivo da teoria de eventos proposta por Goldman e Kim. Infelizmente, porém, quando aplicada ao caso específico dos eventos mentais, a teoria gera duas dificuldades sérias. Consideremos, por exemplo, o caso do objeto fenomenal da *afterimage*. Se fizermos da *afterimage* um evento e se eventos são entidades espaço-temporalmente localizadas, então temos de poder localizar o evento da *afterimage*. O candidato mais plausível para a localização, neste caso, é o cérebro. A noção de eventos proposta exige ainda que identifiquemos também o objeto que instancia a propriedade que constitui o evento da *afterimage*. Este objeto só pode ser o sujeito experiencial da *afterimage*. Quando eu sou este sujeito experiencial, o objeto só pode ser o que é designado pelo uso da expressão “eu”, pois sou *eu* que tenho uma *afterimage*. É difícil conceber um modelo que satisfaça simultaneamente todas estas exigências. Ainda que meus estados mentais (minhas experiências) estejam no cérebro, eu não estou no meu cérebro. Além disso,

tampouco sou ou posso ser considerado como parte constituinte da *afterimage* que tenho. É para escapar de tais dificuldades que Tye abandona a teoria de eventos proposta por Goldman e Kim e adota a noção de evento fornecida por Davidson, a saber: eventos como entidades particulares ontologicamente básicas (não-analisáveis) que estão em relações causais. Isto implica, porém, que a solução imediata para o problema do pertencimento necessário, fornecida automaticamente pela teoria de Goldman e Kim, não está mais acessível. A ela, Tye tem de chegar por outra via, ainda não descrita.

Com base no que foi dito até agora, a teoria proposta por Tye se expõe à crítica em três níveis distintos. Podemos, em primeiro lugar, questionar se as propriedades *sui generis* de objetos fenomenais se aplicam igualmente (*no mesmo sentido*) a eventos. Tye assume, por exemplo, que eventos têm sempre um “sujeito” ao qual pertence (necessariamente?) uma propriedade. Mas qual é o “sujeito” do evento “flash de luz” ou da “partida de futebol”? Existe um “sujeito” constituinte do evento designado pela expressão “está chovendo”? Em segundo lugar, podemos recusar como inadequada a concepção davidsoniana de evento adotada por Tye. Esta dá às propriedades um *status* ontológico secundário. A realidade é feita de eventos não-analisáveis. As propriedades têm um caráter derivado. E, por fim, mesmo depois de ter aceitado a transposição das propriedades *sui generis* do âmbito dos objetos fenomenais para o dos eventos e a concepção de evento proposta, podemos indagar em que se distingue efetivamente a alternativa sugerida por Tye, a teoria dos objetos fenomenais como eventos, da já mencionada teoria adverbial eliminativista. Como comentarei mais adiante, parece ainda plausível supor que a teoria de Tye enfrenta os mesmos problemas gerados pela teoria adverbial. A isto Tye replicaria com a afirmação que sua teoria oferece, em contrapartida, a vantagem de não se comprometer em negar a existência de objetos fenomenais, mas simplesmente de identificá-los a eventos; não introduzir advérbios artificiais e não deslocar o problema dos objetos fenomenais para a experiência.

Mais especificamente, Tye trata de três objeções potenciais à sua teoria: (1) a identificação contraintuitiva do processo da experiência com o que é objeto da mesma, isto é, a identificação do “ter uma *afterimage*” (o evento) com o objeto “tido” neste processo (a *afterimage*); (2) a suposição de que a experiência (o processo) pode ser descrita com predicados que se aplicam, na nossa linguagem ordinária, aos seus objetos—o processo de ter uma *afterimage*, por exemplo, tem de poder ser azul como a *afterimage* “tida” neste processo; o evento de sentir uma dor tem de poder ser intenso como a dor sentida neste evento, etc.; por fim (3) a teoria de Tye parece partilhar com a teoria adverbial da experiência uma solução implausível ao problema de experiências com múltiplos objetos fenomenais; em particular, ter três *blue afterimages* significaria “ser sujeito de três experiências visuais,” o que certamente contraria a visão comum de que só há *uma* experiência neste caso, embora com *três* objetos distintos. Para Tye, a solução destes problemas requer uma investigação da suposta natureza intencional dos eventos ou estados fenomenais.

Tye desenvolve então uma teoria dos eventos ou estados fenomenais como intencionais, embora não-conceituais. Tais estados podem acompanhar outros estados intencionais de conteúdo proposicional. Eles podem servir de *background* para processos cognitivos envolvendo estados intencionais de conteúdo proposicional. O decisivo, porém, é que os estados fenomenais intencionais *não* são eles mesmos conceituais. Sensações corporais como, por exemplo, dor de dente e orgasmo são representacionais no sentido em que refletem causalmente, *em circunstâncias normais*, as alterações em partes específicas do corpo. Estas alterações formam o conteúdo representacional das sensações corporais. Tais representações (i) são o *output* de receptores nervosos respondendo a estímulos diversos no nosso corpo e (ii) integram o *input* dos processos cognitivos em que são, pela primeira vez, reveladas cognitivamente ao sujeito das representações (Tye diz, nestes casos, que mediante a ação de processos cognitivos operando sobre as representações não-conceituais o sujeito se torna consciente de (*aware of*) seus estados fenomenais). Tornar-se consciente de (*to become aware of*) suas próprias experiências é um processo cognitivo que subsume, pela primeira vez, a experiência fenomenal a conceitos. Só então o sujeito experiencial adota uma atitude epistêmica (cognitiva) frente a seus estados fenomenais. *Abaixo* deste nível de *awareness*—chamado tradicionalmente de “introspecção”—está a consciência fenomenal (*consciousness*).

Vejamos se Tye estaria, agora, em condições de responder às três objeções já mencionadas. A primeira delas diz respeito à identificação contraintuitiva da experiência (do processo de “experienciar”) com o objeto da experiência. O que há de estranho nesta identificação desaparece quando a experiência é vista como evento. Neste caso, o objeto da experiência (do evento representacional) é o conteúdo intencional da mesma (o que é representado). As *afterimages* ou dores *não* são conteúdos neste sentido amplo, mas sim o que é experienciado quando determinados estímulos atuam sobre o corpo. Estes estímulos, as alterações em partes específicas do nosso corpo ou suas causas no meio ambiente, são, segundo a teoria de Tye, os objetos representados pelos eventos fenomenais “dor,” “afterimage” etc. E ao identificar o objeto fenomenal “dor” com um evento representacional a teoria *não* confunde a representação com o seu objeto no sentido amplo. A dor *não* é o objeto intencional (representacional) da experiência “dolorosa”. A dor (o evento experiencial) representa para mim alguma alteração relevante no meu corpo. Esta alteração física é então representada para mim como dor.

A segunda objeção, ou seja, a afirmação de que a atribuição de propriedades *sui generis* à experiência forçar-nos-ia a adotar uma linguagem com a qual não estamos familiarizados, ou seja, nos forçaria a atribuir à própria experiência predicados que costumamos atribuir aos objetos da experiência, não parece representar um autêntico problema. Parece estranho dizer que um evento é azul. Parece que com isso estamos transferindo ilicitamente uma propriedade de objetos a entidades onde não faz sentido aplicá-las. Porém, se não estivermos dispostos a aceitar que possamos reportar-nos a eventos azuis, também deveríamos recusar-nos a

falar de *afterimages* azuis. O que afirmamos verdadeiramente quando dizemos que uma *afterimage* é azul é a nossa experiência representa, ainda que falsamente, alguma coisa no mundo como sendo azul. Nossa linguagem estaria, portanto, tão pouco em ordem num caso como no outro.

A terceira crítica merece maior atenção. Parece realmente contraintuitiva a ideia de que ter uma experiência com múltiplos objetos deva ser analisado como um processo composto de múltiplas experiências, cada uma delas com seu objeto próprio. Quando tenho uma *afterimage* vermelha à esquerda de uma *afterimage* verde, por exemplo, não estou tendo a experiência de uma *afterimage* vermelha e a experiência de uma *afterimage* verde, ou seja, duas experiências distintas, mas sim uma experiência onde uma representação de algo vermelho aparece concomitantemente a uma representação de algo verde. Como Tye reconhece: “minha experiência representa que algo vermelho está à esquerda de algo verde” (110). Não estaríamos, assim, postulando várias experiências, mas uma única experiência originada em múltiplos estímulos corporais. Um único evento está sendo vivenciado pelo sujeito. Aqui é difícil evitar a impressão de que Tye está realizando um “truque com palavras.” De fato, ele qualifica o caráter unitário da experiência em questão ao afirmar, imediatamente após esta passagem, que se trata de uma experiência “composta” tendo como componentes uma experiência de algo verde e uma experiência de algo vermelho. Ambas as experiências são da mesma modalidade, ou seja, experiências visuais e atribuídas ao mesmo sujeito. Isto nos leva de volta, porém, ao mesmo ponto a que tínhamos sido conduzidos pela teoria adverbial, para a qual Tye pretendia formular uma alternativa genuína.

No início desta exposição, mencionei a perspectividade como uma das propriedades características de objetos fenomenais. Trata-se da suposição de que, ao contrário dos estados físicos em geral, alguns dos nossos estados só podem ser inteiramente compreendidos por aqueles que estão no próprio estado. A compreensão da natureza destes estados requer, portanto, que se adote frente a eles o ponto de vista, ou a perspectiva, de quem os vivencia. De acordo com o famoso exemplo de Nagel (1979), podemos saber tudo acerca de um morcego, compreender seu modo de orientação no mundo, o completo funcionamento de seu sistema perceptivo baseado na emissão e detecção de ondas sonoras e tudo mais que caia sobre as leis do mundo físico. Jamais saberemos, no entanto, o que é *ser como um morcego*, ou seja, como é ser como *aquele que vivencia* tais processos perceptivos. O que quer signifique *saber* neste sentido, parece ser algo que não pode ser concebido como já tendo sido dado pela nossa compreensão física do mundo. Para designar esta propriedade aparentemente opaca ao conhecimento do mundo físico, Nagel utiliza a expressão “*what it is like (to be)*.” Haveria, então, algo novo que realmente só somos capazes de apreender quando *vivenciamos* certos estados? Haveria, por fim, objetos ou fatos irreduzíveis aos já contidos em um conhecimento pleno do mundo físico? Vejamos, agora, como a teoria de Tye supõe responder a esta questão e, com isso, afastar, mais uma vez, o fantasma do dualismo.

A resposta de Tye surge a partir de uma reconstrução do chamado argumento do conhecimento ou, em outras palavras, do experimento de Mary (F. Jackson 1982). Mary é uma cientista brilhante que sabe tudo acerca do mundo físico. Mary nasceu, cresceu e aprendeu tudo o que sabe em um quarto preto e branco, de tal forma que embora conheça todas as propriedades dos objetos físicos, inclusive as cores, jamais teve, ela mesma, a experiência de ver cores. Para o experimento é importante conceder que Mary vê inclusive a si mesma em preto e branco e que ver preto e branco já a torna apta a compreender a categoria de “cores”. Um dia, Mary é retirada do quarto e vê pela primeira vez tomates vermelhos.

Pergunta-se: Mary aprendeu alguma coisa nova ao deixar o quarto? Se nossa resposta for—como parece intuitivamente ser—“sim,” então como conciliar a premissa de que ela já possui todo o conhecimento possível acerca do mundo físico com este novo saber? Aparentemente estaríamos diante de duas alternativas: (1) Aceitamos que Mary aprendeu algo novo e, posto que ela já sabia tudo acerca dos objetos e leis do mundo físico, teríamos, então, de admitir a existência de fatos ou objetos não-físicos. Ou (2) simplesmente negamos que Mary tenha apreendido algo novo. A primeira alternativa é claramente dualista. Ela admite um desenvolvimento *interacionista* ou um desenvolvimento *epifenomenalista*. No primeiro caso, as propriedades experienciais que formam o conteúdo do novo conhecimento de Mary, as propriedades do tipo “*what it is like*,” desempenham um papel causal no comportamento posterior de Mary e podem ser invocadas na explicação de seus novos hábitos e capacidades discriminatórias. No segundo caso, o “*what it is like*” é real, porém causalmente inerte. Esta foi a solução endossada por Jackson na formulação original do caso de Mary. A segunda alternativa é *eliminativista*. Ela nega simplesmente que existam propriedades experienciais do tipo “*what it is like*” e com isso também a perspectividade essencial de alguns estados do organismo. Devemos supor que tais alternativas esgotam o âmbito de soluções possíveis para o problema de Mary? Ou existe uma opção ainda não explorada? Como compatibilizar nossas convicções fisicalistas com a nossa mais trivial intuição de que Mary efetivamente vivenciou algo novo ao deixar seu mundo preto e branco?

De acordo com a descrição do experimento, Mary sabe, antes de sair do quarto, o que são cores e já possui uma noção de vermelho aplicável a tomates maduros, hidrantes, sangue, cabines telefônicas inglesas etc. O que ela não pode saber é o que é *ser como alguém que vê vermelho*, posto que isto suporia passar pela experiência de ver vermelho. Ao sair do quarto Mary tem esta experiência. Devemos considerar o saber o que é estar no lugar de alguém que vê vermelho como um saber adicional? De acordo com a teoria de Tye, algo foi adicionado ao conceito que ela já possuía de vermelho, a saber: seu aspecto fenomenal. Com a nova experiência de Mary, novos elementos são integrados ao *input* dos processos cognitivos que levam Mary ao reconhecimento de tomates maduros. Há, por assim dizer, um novo conceito, o conceito fenomenal de vermelho, adquirido como *output* de processos iniciados com alterações registradas por receptores corporais em respostas a estímulos aos

quais só agora Mary é exposta. Este novo conceito a habilita a fazer discriminações no espaço das cores. Mary pode, agora, dirigir sua atenção para as cores como uma forma mais eficaz de selecionar tomates maduros. Não se trata, portanto, necessariamente da apreensão de propriedades ou fatos não-físicos, mas da aquisição de um novo conceito com base em representações não-conceituais que passam a integrar os processos cognitivos de Mary e a habilitam a se relacionar de novas formas com seu mundo. Seu novo conceito tem a mesma extensão dos conceitos físicos que Mary já possuía. Em outras palavras, o conceito fenomenal de vermelho, recém adquirido por Mary, se refere a uma propriedade que pode ser completamente identificada pelos conceitos físicos que Mary anteriormente associava à cor vermelha. Esta propriedade, segundo Tye, é PANIC. Há, portanto, uma nova forma, perspectiva, de se referir a uma propriedade física. Num certo sentido há apenas um fato que constitui, antes e depois, o conteúdo do conhecimento de Mary. Em outro sentido, há, porém, um fato novo que inclui a nova forma de acesso perspectivo ao conteúdo da experiência de Mary. O novo fato, o fato desconhecido anteriormente por Mary, é constituído por conceitos fenomenais.

À interpretação de Tye, podemos acrescentar a observação de que a experiência de Mary, agora acompanhada de novas representações, fornece a ela um novo *know how*. Com isso, podemos supor que Mary se tornou capaz de responder de forma mais hábil e adequada a inúmeras situações às quais podemos submetê-las. Entre outras coisas, seu comportamento certamente já não será o mesmo ao escolher as roupas que pretende vestir, ou ao eleger as cores de sua nova casa. Podemos ainda supor que esses novos elementos a habilitam a responder de forma direita, mais rápida e eficaz a certas situações.

Imaginemos, por exemplo, a seguinte situação: Mary acaba de sair de casa, quando vê caminhar lentamente em sua direção uma pessoa cujo corpo está coberto por um líquido espesso, identificado através da sua cor, vermelho, como sendo sangue. Mary se precipita sobre a pessoa e estendendo-lhe às mãos. Podemos supor que as impressões visuais do sangue foram imediatamente registradas por Mary, fornecendo, assim, uma base de resposta imediata à situação em questão. Ao relatar o que aconteceu, ou ao justificar sua atitude frente a terceiros, Mary poderá afirmar ter identificado prontamente a situação como a de uma pessoa ferida em busca de socorro e que, ao estender a mão, agiu com base neste suposto saber. Nos dois momentos, o da experiência propriamente dita e o do relato da experiência, trata-se, salvo enganos, de uma pessoa ferida. O que distingue os dois casos não é, portanto, o conteúdo do que é percebido, ou seja, o conteúdo da experiência, mas o modo como ela é registrada por Mary. As impressões visuais do sangue provocam em Mary uma reação imediata que não precisa estar acompanhada de qualquer processo reflexivo. A narrativa do evento, por sua vez, supõe um conteúdo proposicional para o qual Mary está dirigida, ou seja, um conteúdo proposicional intencional que, neste caso específico, tem como fonte um conteúdo intencional não-proposicional, a saber: as impressões visuais registradas por Mary.

Podemos nos perguntar o que haveria acontecido, neste caso, se Mary ainda não tivesse passado pela experiência de ver vermelho. Recorrendo à sua base de dados, ou seja, a todo o seu conhecimento acerca dos objetos e das leis do mundo físico, Mary certamente teria chegado à conclusão de que a pessoa diante de si está ferida e finalmente, reagiria de forma adequada. Contudo, a compreensão de sua própria experiência e seu relato a terceiros não nos reportaria a impressões visuais de cores, mas a outros conteúdos cognitivos. Mary teria, por assim dizer, uma via indireta de acesso a certas informações, o que certamente comprometeria sua reação. Neste sentido, ao sair do quarto, Mary passa a poder contar com uma nova via de acesso a informações, uma via direta, que poderá induzi-la a novas atitudes diante dos mesmos fatos.

Uma solução semelhante à de Tye é proposta por John Perry em seu último livro *Knowledge, Possibility and Consciousness*. Perry reconstrói o caso de Mary com o objetivo de defender o fisicalismo. Em particular, Perry procura mostrar que o *know how* adquirido por Mary é uma forma de *know that*.

Para Perry, o que gera realmente a dificuldade no argumento do conhecimento não é o fisicalismo, mas uma concepção inadequada, embora comum, do que seja conhecimento. Segundo Perry nossa análise do conhecimento não deveria levar em conta apenas o conteúdo estritamente referencial, mas também o conteúdo reflexivo. Os estados epistêmicos de Mary, antes e depois de sair do quarto, podem ter o mesmo conteúdo referencial estrito, especificado exclusivamente em termos dos referentes externos dos vários itens que compõem o estado mental de Mary. Seus conteúdos reflexivos, no entanto, são distintos. O conteúdo reflexivo é aquele em cuja especificação é feita uma referência ao próprio veículo do conhecimento, às estruturas conceituais que formam o estado mental de Mary. O erro na interpretação antifisicalista do caso de Mary estaria em restringir o conteúdo cognitivo de Mary às condições de verdade impostas aos objetos externos, ou seja, ao conteúdo estritamente referencial. Como em outros casos não controversos explorados por Perry, somente o conteúdo reflexivo forneceria uma explicação razoável da relação dinâmica entre os estados cognitivos de Mary e suas ações. Neste sentido, o novo conteúdo reflexivo justificaria o *know how* adquirido por Mary. Para Perry é, portanto, fundamental que este novo *know how* só possa ser compreendido sobre a base de um novo *know that* com conteúdo analisável em termos de condições de verdade impostas também às estruturas conceituais do estado mental instanciado quando Mary tem sua primeira experiência de cores.

Embora a distinção entre conteúdo reflexivo e conteúdo estritamente referencial forneça a chave interpretativa para a elucidação do caso específico de Mary, bem como de outros casos que envolvem uso de dêiticos, isto é, identificadores demonstrativos espaço-temporais e pronomes pessoais, a generalização sugerida por Perry desta distinção a todo tipo de situação epistêmica carece de justificação. Neste ponto, a distinção equivalente sugerida por Tye entre fatos individuados mediante

conceitos fenomenais (*fine-grained, conceptual facts*), por um lado, e fatos identificados com estados de coisas no mundo objetivo, independentemente do modo como tais estados de coisas são concebidos (*coarse-grained, nonconceptual facts*), por outro, é, no mínimo, mais econômica.

A teoria de Tye dá ainda um passo decisivo além de Perry ao sugerir que o conceito fenomenal adquirido por Mary, o conceito que lhe abre novas perspectivas recognicionais, revela a Mary uma propriedade relacional, não-intrínseca. A informação veiculada pelo aspecto subjetivo (pelo *what it is like*) da experiência de Mary, encapsulada na fórmula PANIC, é relativa às alterações no corpo de Mary ou às suas causas regulares no meio ambiente.

A este resultado chegou também Jackson numa reavaliação da sua posição, publicada em 1998. Para Jackson o verdadeiro desafio colocado pelo caso de Mary é explicar por que temos a intuição que Mary aprende alguma coisa que excede o que podia ser deduzido da visão física de como são as coisas. Apenas aparentemente a informação adquirida por Mary é informação sobre propriedades intrínsecas. Como esta informação não é obviamente sobre propriedades *físicas* intrínsecas, o antifisicalista é levado a ver aqui informações sobre propriedades não-físicas. Para Jackson é nesta inferência que devemos localizar o erro da interpretação antifisicalista. O que Mary adquire é, segundo ele, informação sobre propriedades relacionais e funcionais de processos em nós. Jackson afirma, neste ponto, não ter coisa alguma a acrescentar aos argumentos detalhados dos fisicalistas que, antes dele, formularam teorias representacionais da consciência fenomenal (Jackson 1998, p. 78).

O que podem saber os seres não-oniscientes: sobre a resposta de Lynne Baker ao fisicalismo³⁹

Em *Explaining Attitudes*,⁴⁰ Lynne Baker empreende a tarefa de restituir o caráter explicativo causal de nossas atitudes proposicionais, ou seja, dos estados mentais como convicções, desejos, expectativas, enfim, o variado leque de nossos estados psicológicos ou intencionais de conteúdo proposicional. Segundo Baker, ou bem somos capazes de atribuir caráter causal às explicações envolvendo tais estados, ou teremos que deixar de lado grande parte, senão quase a totalidade, de nossas práticas explicativas cotidianas. A atitude adotada por Baker, o “realismo prático”, baseia a causalidade nas práticas cotidianas e científicas bem estabelecidas, se opondo assim ao quadro vigente na filosofia na mente, que apela a uma metafísica dos estados mentais e à sua identificação com estados cerebrais como forma de atribuir-lhes qualquer papel explanatório.

O presente trabalho analisa a proposta geral de Lynne Baker (1). Para tal apresentarei as principais teses e características do realismo prático e de seu par antagonístico, a visão *standard*, como é chamado por Baker. Ademais, o trabalho avalia criticamente o teste dos contrafactuais proposto por Baker como critério para atribuição de poder causal a um determinado tipo de estado mental e a viabilidade aplicativa do mesmo através de sua possibilidade de responder ao problema da suficiência, ou seja, a possibilidade de identificarmos os contrafactuais corretos e suficientes, para a determinação da eficácia causal de eventos mentais.

1. Apresentação do Realismo Prático:

O realismo prático surge, na filosofia da mente, no contexto da investigação do problema da causação mental. Um histórico do problema nos levaria à questão da relação mente/corpo e da possibilidade de “interação” entre ambos, já colocada em destaque por Descartes e seus contemporâneos. A filosofia da mente retoma assim um dos mais importantes problemas da tradição filosófica e traz mais uma vez à cena as soluções outrora apresentadas e as alternativas contemporâneas a elas. Como compreender nossos estados mentais? Se considerarmos que tais estados pertencem ontológica- ou epistemicamente a um sistema não-físico, estaremos adotando a alternativa dualista e teremos, como Descartes, de enfrentar a difícil explicação da possível relação entre um tal sistema, imaterial, não-físico, e o mundo físico. Aceita a

³⁹ Agradeço a Wilson Mendonça pelas discussões que tornaram possíveis este trabalho e por sua cuidadosa leitura e apreciação crítica.

⁴⁰ Baker, L. *Explaining Attitudes*. Cambridge: Cambridge University Press, 1995.

tese do fechamento causal do mundo físico, seria agora como que inevitável pensar a independência e impenetrabilidade de ambos os mundos. Como principal alternativa ao dualismo, impõe-se então a visão *standard*, ou seja, a perspectiva segundo a qual estados mentais são (i) idênticos, (ii) constituídos ou (iii) realizados por estados cerebrais. Os adeptos desta perspectiva constituem a quase totalidade dos investigadores contemporâneos do problema. São eles os fisicalistas em suas diferentes nuances. Em outras palavras, tal perspectiva insere os estados mentais no mundo físico, atendendo ou não a algumas de suas qualidades distintivas. Lynne Baker manifesta por assim dizer um veto quase solitário ao fisicalismo. Para ela há algo de imperdível, imprescindível, que a tradição da visão *standard* não consegue resgatar: nossas práticas explicativas cotidianas bem sucedidas. Não se trata de recair na imaterialidade de nossos estados mentais, mas sim de pensar a insuficiência da explicação fisicalista dos fenômenos mais triviais com que nos confrontamos no nosso dia-a-dia. O que seria um convite para jantar, uma eleição ou mesmo uma sentença de morte se não recorrermos às intenções e crenças a eles subjacentes?⁴¹ Parece haver todo um universo de relações para o qual a atribuição de atitudes proposicionais se torna imprescindível. E o mais importante, neste universo—que abrange a quase totalidade de nossas práticas sociais—nossas explicações têm sido bem sucedidas, ainda que desconheçamos quase que completamente os princípios ou leis físicas que supostamente envolvem seus elementos. Nossas práticas explicativas cotidianas são vistas, assim, como autônomas e não menos reveladoras da realidade do que o conhecimento decorrente das práticas científicas. Baseado em tais experiências, o realismo prático reclama uma nova caracterização de nossas crenças na qual as mesmas não necessitem ser encaradas nem como estados cerebrais, nem como estados imateriais. Exatamente o que podemos compreender como uma crença é o que Baker pretende elucidar com a introdução do teste de controle que será analisado mais adiante.

Ao apresentar-se como uma terceira via de solução para o problema da causalção mental, o realismo prático toma para si a tarefa de mostrar como atitudes proposicionais podem explicar causalmente comportamentos, sem comprometer-se quer com o dualismo, quer com a visão *standard*. Para tal, assumirá, como pressuposto metodológico, a nossa prática cotidiana como fonte de conhecimento. Em outras palavras, o realismo prático, e disto deriva seu nome, assume como ponto de partida uma investigação das nossas práticas cognitivas bem sucedidas. Neste sentido, o que devemos considerar como crenças e desejos será algo que compreendemos observando tais práticas, ou seja, observando os contextos em que tais atitudes são atribuídas. O realismo prático prescinde assim de uma teoria metafísica compreensiva dos estados mentais, mas, segundo Baker, isto não torna menos eficaz suas pretensões metafísicas. Baker insiste que a adoção das práticas

⁴¹ Ver Baker, L. *Explaining Attitudes*, p. 4.

explicativas cotidianas bem sucedidas, sejam elas científicas ou não, como ponto de partida da investigação, deve ser capaz de alterar nossa compreensão dos estados mentais e gerar uma nova compreensão do problema da causação mental e da relação mente/corpo. Como ponto de chegada teríamos, por exemplo, a convicção de que crenças e outras atitudes proposicionais são estados de seres humanos como um todo e não apenas estados de partes de seus corpos, como por exemplo, seus estados cerebrais. Seres humanos ou pessoas são fisicamente constituídos, mas compreender o que seja uma pessoa não se limita a compreender a sua constituição—assim como compreender o que seja o Moisés de Michelangelo não significa compreender o material que o constitui, a fórmula do mármore, mas sim compreender sua inserção no mundo artístico.⁴²

Quais são os exemplos de explicações bem sucedidas que parecem não supor uma metafísica fisicalista, ou a adoção de um fisicalismo *prima facie*, e ainda apontar para a deficiência de uma tal explicação? Explicações fornecidas no âmbito sociológico, econômico, psicológico entre outras, diria Baker. A explicação, por exemplo, de que a queda da bolsa se deu pela redução dos investimentos ou de que a falência da Mesbla foi provocada pelas dívidas da empresa. O que quer que possamos acrescentar acerca da constituição física dos elementos envolvidos parece não contribuir para a explicação já fornecida. Resulta daí um aprendizado que será também transposto para o caso de explicações envolvendo atitudes proposicionais: ainda que possamos reconhecer uma constituição física, nestes casos, ela não contribui para a explicação intencional, psicológica ou não. Ainda que nossas atitudes proposicionais sejam constituídas por estados cerebrais, tal constituição não possui, em certos casos, qualquer poder explanatório. Como o recurso à constituição física poderia tornar mais eficaz a explicação psicológica cotidiana de que João pegou o guarda-chuva porque acreditava que estava chovendo e, além disso, porque não desejava se molhar? No nosso dia-a-dia assumimos a tese de que atitudes proposicionais explicam comportamentos e, quase sempre, nada sabemos acerca das leis estritas que regem nossa constituição física.

É claro que podemos levantar suspeitas quanto aos exemplos fornecidos por Baker e sua interpretação destes exemplos como casos de explicação causal. Podemos suspeitar, por exemplo, (i) que em alguns casos haja entre o *explanandum* e o *explanans* uma relação definicional, isto é, não-causal ou (ii) que haja alguma circularidade inerente a qualquer explicação intencional. Baker não deixe de ser sensível a este tipo de objeção. Quanto ao primeiro tipo de dificuldade, teríamos que analisar caso a caso. Não haveria uma resposta geral. É possível que alguns dos exemplos fornecidos pela autora⁴³ sejam, no nosso entender, casos de relações definicionais, mas a questão poderá ser apenas achar os exemplos certos. Quanto à

⁴² Ver Baker, L. *Persons and Bodies*.

⁴³ Ver Baker, L. *Explaining Attitudes*, p. 100.

segunda suspeita, esta recairia para Baker também sobre qualquer explicação científica, já que fornecer explicações é, antes de tudo, uma atitude proposicional. Não pretendo me deter neste ponto, pois a apresentação de casos de explicações aceitas no nosso cotidiano parece não passar de uma apresentação de tais casos. Ou seja, ainda não diz coisa alguma acerca de como cotidianamente verificamos ou validamos tais explicações. O que importa aqui é que também nas explicações cotidianas precisamos separar explicações “corretas” ou bem sucedidas de explicações “incorretas” ou mal sucedidas. É por isso que o realismo prático não poderá deixar de nos apresentar uma caracterização mínima do que seja uma explicação causal e um critério para sua identificação. Mas antes de chegar a isto, gostaria de apresentar brevemente as principais características das posições compreendidas sob o título de visão *standard*.

2. Visão *standard*:

Em linhas gerais, os defensores da visão *standard* são aqueles que partilham a tese segundo a qual ter um determinado estado mental, em particular, uma crença, é estar num determinado estado cerebral. As dificuldades apresentadas pela identificação de uma crença a um estado cerebral, ou seja, de um estado mental a um estado físico, geram variações significativas entre os teóricos desta perspectiva. Um primeiro grupo é constituído pelos defensores da chamadas “teorias de identidade de tipos” (*type-identity*), onde a identidade em questão está entre tipos de crenças, por um lado, e tipos de estados cerebrais, por outro: “para qualquer crença p , há um estado cerebral de tipo N , tal que para qualquer x , x acredita que p se e somente se x está em N .” Nessa perspectiva, duas pessoas que partilham uma mesma crença estão em um mesmo estado cerebral. A força de tal posição pode ser abrandada por uma discriminação progressiva dos universos equiparados. Ou seja, pode-se, por exemplo, acrescentar: “em seres humanos ‘normais’, ter uma crença p é estar em um estado cerebral N ”.

Contudo, mesmo com maiores refinamentos, a identificação de um tipo de crença a um tipo de estado cerebral parece não dar conta de casos intuitivos em que aparentemente estados físicos diversos realizam um mesmo estado mental. Seres humanos e robôs parecem capazes de realizar os mesmos cálculos matemáticos ainda que em seus “cérebros” diferentes substâncias estejam atuando. Em outras palavras, crenças particulares ou realizações locais de propriedades mentais podem ser idênticas a realizações locais de tipos físicos diversos. Este problema, entendido como o problema da realização múltipla, gera como variação as “teorias de identidade de instanciações” (*token-identity*). De acordo com esta perspectiva, toda instanciação de uma crença particular é simultaneamente a instanciação de um tipo físico qualquer, porém não necessariamente de um mesmo tipo físico. A identidade é estabelecida, aqui, entre instanciações locais de propriedades mentais e instanciações locais de estados cerebrais particulares. Falantes de diferentes línguas, por exemplo, poderão ter tipos de estados cerebrais diferentes realizando a mesma crença, mas para o

falante de cada língua haverá uma realização física particular que expresse a crença em questão.

Uma suposta alternativa à identificação mais forte ou mais franca entre o mental e o físico é oferecida através da substituição da relação de identidade pela relação de constituição. Crenças são constituídas por estados cerebrais, dirão os adeptos das “teorias da constituição”. Se na ótica das teorias da identidade, ao menos instanciações particulares de estados mentais deveriam ser idênticas a instanciações de estados físicos, na perspectiva da constituição, basta que haja uma identidade “molecular”, mesmo que em distintas realizações locais de estados cerebrais.

Uma quarta alternativa, a perspectiva funcionalista, desloca o eixo identificatório e, como diz o próprio nome, volta-se para a *função* realizada por um determinado estado cerebral. Estados mentais particulares são assim identificados à função realizada por estados cerebrais ou ao estado cerebral que realiza esta função. O que torna um estado cerebral uma crença de conteúdo p é a sua relação com determinados estímulos sensoriais, com outros estados internos e com o comportamento. Um estado cerebral é, por exemplo, a crença de que a banana alimenta, quando, em virtude da mesma, o agente se dirige à banana, a leva até a boca e come, quando tem fome.

A sustentabilidade das distinções entre as quatro perspectivas até agora mencionadas não será aqui analisada.⁴⁴ É possível que o enfraquecimento do laço identificatório crie dificuldades sobressalentes para os que o defendem. A manutenção de um algo além do que foi identificado ao estado físico pode gerar a suspeita de um excedente inócuo, sem função e sem poder explicativo/causal. Por outro lado, quando as crenças são totalmente identificadas aos estados cerebrais, parece inevitável admitir que só restam os estados cerebrais e que o algo além, supostamente contido nas crenças, não passe de uma ficção. Assim concluiriam os materialistas eliminativistas. Estes são os mais radicais entre os fisicalistas—mas, para Baker, não necessariamente os mais perigosos ou mais equivocados. Adeptos da visão *standard* incorreriam nos mesmos erros básicos, sejam eles radicais ou não.

Mas quais seriam os equívocos metafísicos concretos da visão *standard*? Há no diagnóstico de Baker pelo menos dois equívocos fundamentais. O primeiro consistiria em um erro categorial: o de confundir relações de superveniência com relações de constituição. Relações de superveniência se dão entre propriedades enquanto relações de constituição se dão entre coisas. Dizemos que um conjunto de propriedades mentais m supervêm ao conjunto de propriedades físicas f , e que um

⁴⁴ Sobre isto ver Kim, J. *Mind in a Physical World. An Essay on the Mind-Body Problem and*

Mental Causation. Cambridge, Mass.: MIT Press, 1998.

determinado objeto x é constituído por um determinado objeto material y . Para Baker, é possível negar a tese da superveniência e aceitar a da constituição, pois a constituição de um determinado objeto não abarcaria tudo que podemos dizer ou saber acerca do mesmo. Haveria, assim, lugar para outro tipo de explicação, em muitos casos, mais completo e adequado. Quem, por exemplo, fala da nona sinfonia apenas a partir de sua constituição, parece não entender sequer do que se trata. É como alguém que descreve um Mondrian apontando para as propriedades das tintas e da tela que o constituem. Há elementos necessariamente relacionais na caracterização de certos objetos e/ou na compreensão de certas situações.

O outro grande erro estaria em negar toda esta estrutura explicativa que apela a um universo de relações e convenções humanas e apostar em uma explicação física mais profunda que viesse a suplantá-la. No diagnóstico da autora, devemos reconhecer que se tratam de explicações diferentes que respondem a diferentes questões. Caso sejamos médicos podemos estar realmente interessados em saber a causa física (microfísica) da cegueira de João. Mas um leigo ficaria mais satisfeito ao saber se foi catarata, um problema congênito ou um acidente de trabalho que a provocou. Aliás, em um tribunal, por exemplo, seria só isso o que interessaria para responsabilizar ou não a empresa, para pagar ou não o seguro devido. Não há, portanto, graus de profundidade, nem respostas erradas a serem suplantadas por respostas verdadeiras, mas respostas diferentes para perguntas diferentes, diria Baker.⁴⁵

3. Sobre as explicações causais em geral e o teste de controle:

Não poderíamos agora concordar com Baker e apenas acrescentar que a distinção entre os dois tipos de explicações mencionados estaria justamente no fato de que de um lado teríamos explicações causais e, de outro, meros esclarecimentos “conceituais”? Podemos dar mais ênfase ao aspecto esclarecedor e passarmos a falar de razões em contraposição a causas e, como bons wittgensteinianos, invocar a impossibilidade ou o *nonsense* de falar de causas ao invés de razões em certos contextos explicativos. Esta alternativa, aparentemente banida das discussões filosóficas depois de Davidson, parece agora recuperar suas forças. *Practical Reasons* de J. Dancy é um bom exemplo disto.⁴⁶ Lynne Baker, contudo, não aceita tal opção. Trata-se de reaver o caráter explicativo causal de enunciados que se referem a atitudes proposicionais.

Insistindo no caráter causal de tais explicações, a autora estabelece condições mínimas para que uma explicação seja efetivamente deste tipo. Em primeiro lugar tais explicações deverão fornecer informações acerca da ocorrência de um fenômeno

⁴⁵ Ver Baker, L. *Explaining Attitudes*, p. 134.

⁴⁶ Ver Dancy, J. *Practical Reality*. Oxford: Oxford University Press, 2000.

particular. Em segundo, deverão mencionar propriedades explanatórias que (i) sejam instanciadas antes da ocorrência do fenômeno explicado e que (ii) não estejam relacionadas ao fenômeno explicado por definição. Em terceiro, devem suportar contrafactuais. A primeira condição parece comum a qualquer tipo de explicação. Neste sentido, satisfazê-la não é um privilégio de uma explicação de tipo causal. A segunda condição parece tentar evitar o risco de circularidade ou petição de princípio, por um lado e mero formalismo e vacuidade conceitual, por outro. Considerando que definições não são normalmente aceitas como explicações de tipo algum, mais uma vez não haveria ainda, aqui, qualquer peculiaridade das explicações de tipo causal. A ênfase da argumentação de Baker recairá, portanto, sobre a quarta condição e dela será retirado o teste proposto pela autora para o reconhecimento de uma relação causal. Trata-se do teste de controle:

(CT) Uma ocorrência de F em C explica causalmente uma ocorrência de G em C se (i) se F não houvesse ocorrido em C , então G não haveria ocorrido em C e se (ii) dado que F ocorreu em C , a ocorrência de G era inevitável.⁴⁷

A primeira condição afirma que a ocorrência de F em C é contrafactualmente necessária para a ocorrência de G em C . A segunda condição afirma que a ocorrência de F em C é suficiente para a ocorrência de G em C .

Aplicando agora o teste ao exemplo da queda da bolsa de valores, teríamos então o seguinte resultado. A redução dos investimentos, no dia 28 de agosto do ano passado, explica causalmente a queda da bolsa de valores neste mesmo dia. Pois, se a redução dos investimentos não houvesse ocorrido, então a queda da bolsa não haveria ocorrido; e dado que a redução dos investimentos ocorreu a ocorrência da queda era inevitável. Da mesma maneira, seria também explicada a falência da Mesbla. Se a empresa não tivesse tantas dívidas, não teria falido. Dada a existência das dívidas a falência era inevitável.

As vantagens de CT são, para Baker, evidentes. Em primeiro lugar o teste contempla a ideia de que causas interferem em seus efeitos. Em segundo lugar, o teste não distingue explicações psicológicas de outras explicações, podendo, portanto, ser aplicado a qualquer tipo de explicação. Em terceiro, não supõe de antemão uma concepção da explicação causal como uma afirmação acerca dos estados internos. E, finalmente, embora não pressuponha o materialismo, é compatível com ele.⁴⁸

Aparentemente, o teste parece restituir o caráter causal de muitas das explicações fornecidas no âmbito da sociologia, economia, psicologia e no nosso dia-a-dia. Porém, ainda considerando os dois exemplos acima, sabemos que, sob o ponto de vista econômico, a segunda condição poderia não ser satisfeita. O governo, o FMI

⁴⁷ Ver Baker, L. *Explaining Attitudes*, p. 122.

⁴⁸ Ver Baker, L. *Explaining Attitudes*, p. 125.

ou alguma outra instância poderia injetar capital como forma de evitar a quebra do sistema financeiro ou a falência de uma empresa devedora. Isto faria, é claro, com que o fato a ser explicado não ocorresse e, como consequência, não estaríamos buscando explicar coisa alguma. O que quer dizer então que a segunda condição pode ser não-satisfeita? Quer dizer que ela só faz sentido se pudermos apresentar a situação *C* de tal forma que ela já exclua os elementos que contribuem para não-satisfação da segunda condição. Ou seja, *C* teria que conter dados como: sem a ajuda do governo, do FMI ou de qualquer outra fonte. Baker, contudo, poderia defender que a introdução destes elementos na descrição da situação não é problemática, por se tratar apenas de um teste para validar ou não as explicações efetivamente fornecidas de atos que efetivamente ocorreram. Quanto mais soubermos acerca das circunstâncias que envolvem o fenômeno a ser explicado, mais aptos estaremos para aplicar o teste, validar e recusar algumas das explicações fornecidas.

Mas até que ponto podemos realmente saber o suficiente para aplicar o teste de forma adequada? Sem um conhecimento satisfatório da situação que nos permita identificar os contrafactuais certos, o teste se torna vulnerável. Para analisar especificamente este problema, pretendo agora investigar a aplicação do teste dos contrafactuais, proposta por Baker, agora, para verificar a eficácia causal de estados intencionais.

4. O problema da suficiência:

Se, como os demais estados intencionais, crenças não são nem estados cerebrais, como propõe a visão *standard*, nem algo misterioso, imaterial, como parecem supor os dualistas, o que são afinal? Já vimos que é peculiar ao realismo prático tomar qualquer estado intencional como algo que se atribui não a uma parte de uma pessoa, a seu cérebro, por exemplo, mas a uma pessoa na sua totalidade. Em outras palavras, a atribuição de um estado intencional específico é feita a um ser que age, fala e pensa de acordo com enunciados que podem ser especificados intencionalmente. O recurso a tais enunciados permitirá validar ou não a atribuição de um estado mental específico ao agente. Seguindo então a proposta de Baker, a verdade ou falsidade da atribuição de uma crença de conteúdo *p* a um sujeito *S*, dependerá da existência de contrafactuais relevantes válidos para *S*. Nas palavras da autora: “The Practical Realist holds this: *S* believes that *p* if and only if, in *S*’s context, there are counterfactuals nonvacuously true of *S*”.⁴⁹

O realismo prático assume, assim, o teste dos contrafactuais não apenas como critério para o reconhecimento de explicações de tipo causal, mas para a própria validação da atribuição de estados intencionais a um sujeito. De acordo com (CT), segue-se, então, que condições contrafactuais relevantes determinam os estados mentais em que o agente se encontra. Tal teoria se estende ao que o sujeito faria, diria

⁴⁹ Baker, L. *Explaining Attitudes*, p. 158.

ou pensaria em circunstâncias diferentes da circunstância real. Nosso problema consistirá, mais uma vez, na especificação das condições contrafactuais suficientes para a atribuição dos estados intencionais em questão. Vejamos um exemplo. Pensemos nas condições necessárias para que seja atribuída a um sujeito *S* (Smith) a crença de conteúdo *p* (FHC é o presidente do Brasil). Seguindo a especificação das condições apresentadas acima (CT) teríamos:

- (i) Se Smith fosse perguntado sobre se FHC é o presidente do Brasil, se Smith compreendesse a pergunta e, além disso, fosse cooperativo, Smith responderia “sim”.
- (ii) Se Smith não acreditasse que FHC é o presidente do Brasil, então ele não responderia “sim”.

Infelizmente, também aqui sabemos que a segunda condição pode não ser satisfeita. Mesmo se Smith não acreditar que FHC é o presidente do Brasil, poderá ainda responder que “sim”. Smith pode ter o hábito de responder afirmativamente a todas as perguntas que lhe são feitas, independentemente de suas demais convicções. Poderá, por exemplo, ter sido educado de forma a responder afirmativamente tudo que lhe seja indagado por um superior. Aqui poderíamos tecer uma série de considerações a respeito da personalidade de Smith e das circunstâncias que o envolvem. Tais considerações poderiam compor um quadro onde o reconhecimento de que Smith compreendeu a pergunta e foi cooperativo não se tornasse incompatível com uma resposta afirmativa, mesmo que Smith pudesse não acreditar ser FHC o presidente do Brasil.

Em um caso similar, onde não por coincidência o sujeito também é Smith, Baker reconhece que as condições contrafactuais expressas nos termos acima seriam realmente insuficientes para determinar o estado mental de Smith. O que Baker, contudo, não reconhece é que este seja um problema capaz de abalar o teste proposto pelo realismo prático. Trata-se, no seu entender, de identificar corretamente os contrafactuais adequados. Ou seja, uma discriminação exaustiva de todos os elementos relevantes, para cada caso, permitir-nos-ia a correta identificação dos contrafactuais relevantes. Nas palavras de Baker:

The Practical Realist’s solution to the sufficiency problem is to concede that the pair of counterfactuals, (i) and (ii), are indeed insufficient for Smith’s believing (that *p*). But it does not follow that there are no counterfactuals true of Smith that are... sufficient for Smith to believe (that *p*). What follows is that we may be mistaken that we are in possession of sufficient conditions for Smith’s believing. We may even be justified (though wrong) to think that (i) and (ii) are sufficient for Smith to believe (that *p*). *What we are missing is not information about Smith’s beliefs or change belief, but rather other relevant information.*⁵⁰

⁵⁰ Baker, L. *Explaining Attitudes*, p. 165f.

Um ser onisciente poderia ter todas as informações relevantes e assim determinar adequadamente o estado mental de Smith.⁵¹

Sob a pressuposição de que a aplicação adequada do teste supõe um conhecimento de todos os fatos envolvidos na situação analisada, a teoria de Baker parece imune a críticas. Contudo, a aplicabilidade da teoria torna-se restrita quando nós, seres não-oniscientes, tentamos determinar os verdadeiros antecedentes causais das ações de outros seres humanos. O aspecto prático da proposta parece ceder lugar a uma idealização injustificada. O teste, assim como a proposta do realismo prático em geral, parecia basear sua relevância na possibilidade de compatibilizar de forma mais satisfatória nossas práticas explicativas cotidianas. Agora, no entanto, acabamos por concluir que os agentes destas mesmas práticas cotidianas não satisfazem as condições de aplicabilidade do teste. Curiosamente o teste só pode ser adequadamente aplicado por um ser não-real, onisciente, como o proposto por Descartes para resolver nossos problemas da interação entre o físico e o mental.

Ao retirar do nosso alcance as condições necessárias ao adequado reconhecimento de relações causais, Baker perde também a possibilidade de fazer frente ao fisicalismo: entre os fatos que nos são desconhecidos, mas que poderiam ser apreendidos por um ser onisciente, podemos supor que estejam exatamente aqueles antecedentes físicos que determinam causalmente, segundo o fisicalista, todos os eventos reais. O ser onisciente estaria em condições de fornecer a explicação física completa do mundo, que excluiria, como supérflua, toda explicação de ações em termos intencionais. Nós, seres não-oniscientes, devemos, portanto concluir que, pelo menos até o momento, não dispomos de instrumentos suficientes para refutar o fisicalismo. Sob este aspecto, estaríamos, então, mais seguros se defendêssemos um fisicalismo *prima facie*, ou seja, se adotássemos o fisicalismo, até que alguém nos provasse ser outra a opção mais adequada.

Quanto à proposta de Baker, fica a pergunta feita anteriormente: por que não (i) aceitar que haja dois tipos de explicação, (ii) não reclamar o caráter causal para as explicações de tipo intencional, mas (iii) assumir que estas possam ser as mais adequadas, ou seja, pragmaticamente mais bem sucedidas, na grande maioria dos casos que envolvem relações humanas?

⁵¹ Ver Baker, L. *Explaining Attitudes*, p. 166.

Quining the Gap: a consciência fenomenal revisada*

In this article, I will argue that there is no gap between physicalist explanations and phenomenal consciousness, as a matter of principle. Based on an analysis of certain *sui generis* properties of phenomenal consciousness, I aim to show that the apparent explanatory gap between the mental and the physical can be eliminated. More specifically, I want to defend the claim that the so-called *sui generis* qualities of mental events, referred to as *qualia*, are physical properties. These properties would constitute, as Tye puts it, the nonconceptual representational content of our conscious experiences. Moreover, they are the input of physical processes whose output would be intentional states with propositional content.

Em seu artigo “A problemática da consciência e a lacuna explicativa” (Kaufmann 1999) procura abordar o problema da redução da explicação de fenômenos que envolvem aspectos mentais à explicação fisicista dos mesmos. Após colocar o problema com base em uma breve exposição das características peculiares aos dois tipos de fenômenos: fenômenos físicos e mentais, Kaufmann analisa os diversos usos do conceito de consciência, tentando lançar maior clareza sobre o dito *explanandum* da problemática em questão. Neste processo, destaca então o conceito de consciência fenomenal como aquele que reuni as propriedades *sui generis* do mental, ao menos aparentemente, irreduzíveis a uma abordagem fisicista. Numa terceira etapa, analisa criticamente a posição de McGinn e Chalmers. Finalmente, conclui apontando para o impasse a que chegamos diante da insustentável tentativa de demonstrar que a explicação do mental é, por princípio, irreduzível a explicação física e nossas mais familiares intuições de que a explicação fisicista dos fenômenos mentais parece deixar de fora o *essencial* da problemática da consciência.

Neste artigo, pretendo endossar a colocação do problema e o destaque dado ao conceito de consciência fenomenal propostos por Kaufmann, e corroborar a tese de que o problema do gap entre explicações mentais e fisicistas não pode ser encerrado como uma questão de princípio. Com base na análise da consciência fenomenal e das propriedades *sui generis* do mental, pretendo tomar uma posição quanto ao *explanandum* do problema que elimina o aparente fosso explanatório entre o mental e o físico. Mais especificamente, pretendo defender a tese de que as chamadas qualidades *sui generis* dos eventos mentais, reunidas sob o título de *qualia*, se reportam a eventos físicos com características peculiares. Estes constituiriam, tal como Tye sugere, o conteúdo representacional, não-conceitual, do *input* de processos físicos cujo *output* seriam outros estados intencionais de conteúdo proposicional, sendo estes últimos passíveis, ao menos *prima facie*, de uma explicação funcionalista.

Por fisicismo podemos compreender, em linhas gerais, a perspectiva segundo a qual tudo o que existe ou que pode ser conhecido pertence ao mundo físico: ao

domínio constituído pelos objetos e leis de uma microfísica elementar do futuro. Qualquer que seja a definição adotada de “mundo físico,” parece, no entanto, haver um conjunto de propriedades *sui generis* aplicáveis *de fato* a um grupo específico de objetos que, pelo menos à primeira vista, *não podem* ser atribuídas significativamente aos objetos físicos. É neste sentido que pareço ser o sujeito único da minha dor e que a mesma pareça só existir quando referida a mim. Dizemos, por exemplo, que “ninguém mais pode sentir a minha dor,” assim como também acreditamos que dor, tristeza, alegria e demais estados de espírito só existem enquanto são vivenciados ou enquanto pertencem a um sujeito. Minhas experiências perceptivas conscientes, minhas sensações corporais, emoções, paixões e humores parecem reportar-se a mim de tal forma que meu carro e minha roupa, por mais personalizados e relacionados à minha personalidade que eu queira interpretá-los, efetivamente não o fazem. Outra pessoa poderia ter o meu carro e a minha roupa, que, de resto, não deixam de existir tão logo deixem de ser posse de alguém. A “monopolização” de fenômenos mentais experienciais por um sujeito, por um “eu,” parece assim caracterizar uma forma de pertencimento intrínseco alheia aos objetos que estão contingentemente sob meu poder. Não deveríamos, então, distinguir os objetos aos quais tenho um acesso privilegiado, enquanto sou aquele que os *vivencia intrinsecamente*—isto é, os objetos que só podem ser “tidos” e conhecidos na *minha* perspectiva experiencial e que não subsistem a não ser em relação com sujeitos de experiência—de outros objetos aos quais me relaciono simplesmente porque sou aquele que os *tem* contingentemente? Se as propriedades características deste grupo de objetos, isto é, as propriedades da privacidade, do necessário pertencimento a um sujeito e da perspectividade não puderem ser atribuídas a objetos físicos, não deveremos então admitir a existência de entidades não-físicas?

Sob a suposição de que o mundo físico é causalmente fechado—as causas completas de eventos físicos incluem somente outros eventos físicos—estaríamos então impossibilitados de estabelecer relações causais entre o grupo de eventos aos quais supomos pertencer nossas sensações, emoções, paixões e humores, por um lado, e os eventos físicos, por outro. Na melhor das hipóteses, os eventos fenomenais poderiam ser efeitos, mas jamais causas de eventos físicos. Mas estaremos realmente dispostos a admitir que estados fenomenais sejam causalmente irrelevantes?

O modo aparentemente incontornável com que somos tomados por tais convicções e que reagimos a aparente evidência de tais propriedades parece instaurar um fosso entre as explicações fisicistas e o que mais intimamente supomos ocorrer em certos contextos. Seria possível explicar no vocabulário fisicista, ou seja, em termos de uma microfísica elementar, ainda que do futuro, o que é ser o sujeito de determinadas vivências ou retratar este ponto de vista privilegiado que inevitavelmente assumimos diante de nossas sensações, emoções e humores?

Diante deste problema, parecem figurar três alternativas:

1. Aceitar o caráter *sui generis* dos objetos fenomenais e sua irreducibilidade ao domínio do mundo físico;
2. Mostrar que os objetos fenomenais *não* possuem realmente as propriedades *sui generis* mencionadas; ou mostrar que os mesmos não existem e, por conseguinte, que tais propriedades não se aplicam;
3. Reconhecer a existência de objetos fenomenais e de suas propriedades características, negando, porém, que estas não se apliquem também aos objetos físicos.

A primeira alternativa é claramente dualista, ou melhor, reabilita um dualismo ontológico no qual a relação entre o mental e o físico torna-se contingente, arbitrária ou mesmo impossível. Na melhor das hipóteses, seria possível resgatar um epifenomenalismo, onde o mental seria causalmente inerte. Nossas convicções materialistas do século XXI tornam, contudo, difícil compreender o que seriam objetos que não produzem efeito no mundo físico. Em outras palavras, postular a existência de algo parece implicar o reconhecimento de suas relações causais com outros objetos ou eventos físicos.

A segunda alternativa é reducionista ou eliminativista, ou seja: ou bem reduz os objetos fenomenais aos objetos físicos, negando suas características peculiares, ou nega a existência de tais objetos e como consequência o reconhecimento de propriedades *sui generis*. Esta alternativa deixa intacto o fisicismo, porém não nossas práticas cotidianas de atribuição de poderes causais a nossos estados mentais. A terceira alternativa parece compatível com o fisicismo, com um dualismo ou pluralismo de propriedades e com certa compreensão do caráter peculiar de algumas de nossas vivências.

Tanto a segunda quanto a terceira alternativa compartilham o núcleo das intuições fisicistas e a diferença entre ambas é bem mais sutil do que pode nos parecer à primeira vista. Algum tipo de reducionismo ou mesmo de eliminativismo parece inevitável até mesmo à terceira alternativa. A diferença entre ambas consiste apenas na tentativa desta última de conciliar o fisicismo com algumas de nossas mais caras intuições acerca de nossos estados subjetivos. Mas que estados subjetivos especificamente? Ou melhor, que características peculiares aos mesmos parecem razoáveis sustentar junto ao fisicismo, ou seja, sem abandonar o núcleo de nossas convicções fisicistas?

Aqui temos que tornar explícita uma tomada de posição frente ao que seria o *explanandum* do problema. O que parece oferecer resistência a uma redução fisicista não são os estados mentais em geral, mas certas propriedades que supomos pertencer a estados mentais específicos. São, como vimos acima propriedades como: privacidade, necessário pertencimento a um sujeito e perspectividade. Tais propriedades parecem exprimir o caráter subjetivo da experiência, ou seja, se reportam ao que é vivenciado única e exclusivamente na perspectiva do sujeito da experiência. Trata-se do conjunto de vivências para as quais Nagel introduziu a

expressão “what it is like to be” (Nagel 1991). Na perspectiva de Nagel, tais propriedades acompanhariam todo o tipo de evento intencional, ou seja, todo tipo de relação que envolvesse a consciência e seus objetos, constituindo o caráter subjetivo, irreduzível das mesmas. Neste artigo pretendo defender uma posição bem mais restrita frente às mesmas: uma posição que exclui deste âmbito nossos estados mentais como crenças, desejos e intenções cujo conteúdo relevante é uma atitude proposicional. Para justificar tal restrição pretendo mostrar que é possível fornecer, ao menos *prima facie*, uma descrição funcional de nossas atitudes proposicionais e que tal descrição parece não deixar de fora nenhum elemento relevante da compreensão das mesmas. O irreduzível caráter subjetivo da experiência - que estarei aqui compreendendo sobre o título de *qualia* - será, assim encarado como uma qualidade peculiar dos estados mentais relativamente aos quais a perspectiva de quem os vivencia parece acrescentar algo de irreduzível à própria experiência. Com isso pretendo de antemão caracterizar o problema da lacuna explicativa como dirigido especificamente aos estados mentais que parecem não poder ser descritos satisfatoriamente sem uma referência direta àquele que os vivencia ou, em outras palavras, a suas qualidades fenomenais. Antes de passarmos, portanto, ao que considero ser o problema da lacuna explicativa propriamente dita, pretendo expor brevemente o que poderia ser uma explicação funcional de nossas atitudes proposicionais, justificando assim, a restrição proposta.

A perspectiva funcionalista, em geral, caracteriza-se por identificar os estados mentais à função realizada por estados cerebrais ou ao estado cerebral que realiza esta função. O que torna um estado cerebral uma crença de conteúdo *p* é a sua relação com determinados estímulos sensoriais, com outros estados internos e com o comportamento. Neste sentido, os estados mentais são tanto causas internas do comportamento, como efeitos produzidos pelo mundo externo. Nas palavras de Jackson e Braddon-Mitchell:

... a teoria funcionalista da mente especifica estados mentais em termos de três tipos de cláusulas: as cláusulas de entrada que especificam que condições dão origem, tipicamente, a que estados mentais; cláusulas de saída que especificam que estados mentais dão origem, tipicamente, a que repostas comportamentais; e cláusulas de interação que especificam como estados mentais interagem tipicamente. (Braddon-Mitchell e Jackson 1996, p.40)

Para que não haja circularidade na explicação, a caracterização de cada etapa deve ser realizada em termos meramente físicos, ou seja, os estados intencionais que pretendemos descrever em termos físicos não podem ser reintroduzidos na caracterização do *output*. Os *outputs* devem poder ser totalmente descritos em termos de alterações físicas no mundo. A descrição funcional de estados mentais envolve assim uma rede de relações físicas que inclui disposições comportamentais, causas típicas e outros estados mentais, desde que também caracterizados em termos de disposições comportamentais, ou seja, em termos funcionais.

A forma mais usual de exemplificar o funcionamento da teoria é através da construção de modelos, máquinas programadas para realizar um tipo de funcionamento específico. Neste caso, as ilustrações vão desde máquinas mais simples como as que nos oferecem Coca-Cola mediante a introdução de moedas de um determinado valor, até as máquinas conexionistas de tipo PDP (Parallel Distribution Process) ou redes neurais. Parece evidente que a complexidade das funções realizadas por nossos estados mentais exige modelos flexíveis e bastante complexos. Complexos talvez o suficiente para não sermos, ainda, capazes de descrevê-los. Neste sentido a proposta funcionalista parece vulnerável a uma objeção de fato, mas, se estivermos corretos, poderá mostrar que, em princípio, nada impede que todos os aspectos relevantes do mental possam ser descritos em termos físicos.

A mente humana, descrita funcionalmente, poderia, assim, operar como um programa flexível composto de vários módulos. No primeiro módulo estaria um scanner, responsável pela recepção dos *inputs*. A partir daí, podemos imaginar vários módulos entre os quais um módulo avaliador, responsável pela seleção das informações que chegaram à etapa final, ou seja, a produção de um comportamento específico. A peculiaridade deste tipo de programa estaria na sua capacidade de aprender, ou seja, de alterar o produto em função de um novo *input* que incluiria os efeitos produzidos pelo comportamento do agente nas etapas anteriores. Em outras palavras, o *output* gerado promoveria respostas externas que, por sua vez, seriam introduzidas no *input* e avaliadas pelo programa de forma a fornecer um novo resultado. Deste modo, um modelo programado para reconhecer letras, como o que é utilizado com sucesso nos correios, altera, ou melhor, aprimora, sua performance, na medida em que lhe são oferecidas grafias variadas de uma mesma letra e que o programador responde negativa ou positivamente ao seu reconhecimento ou não de uma letra. Tais modelos são capazes de reordenar os seus dados de forma a passar a reconhecer características anteriormente ignoradas. Com base na análise funcional de modelos com programa flexível passaríamos, assim, a atender uma das principais características do modelo humano: sua capacidade de aprendizado. Por esta via, outras características próprias aos humanos, como por exemplo, seu poder de deliberar e intervir no curso de suas ações - algo de que dependem nossas práticas de atribuição de responsabilidade e liberdade a um agente - poderiam também ser resgatadas com sucesso.¹

Supondo, assim, que uma descrição funcional de nossas atitudes proposicionais seja possível e que com tal explicação nada de essencial às mesmas se perca o problema da lacuna explicativa, ou o problema da redução fisicista passa a se concentrar exclusivamente na questão dos *qualia*. Como encaixá-los em uma descrição física do mundo? Ou ainda, a que exatamente tais propriedades se referem?

Se não estivermos dispostos a adotar uma perspectiva reducionista ou eliminativista que recuse de antemão que algo além de nossas próprias convenções linguísticas possa estar na base da crença em tais propriedades e nos objetos aos quais

se reportam, deveremos então ser capazes de fornecer uma descrição física das mesmas, ou em outras palavras, explicar em termos fisicistas como “ser o sujeito de certas vivências” (what it is like to be) pode desempenhar algum papel nas nossas disposições comportamentais, ou na nossa descrição funcional de nossos estados mentais. Tal possibilidade é o que pretendo analisar com base na teoria PANIC proposta por Michael Tye (Tye 1998).

A teoria PANIC é uma teoria dos eventos ou estados fenomenais como intencionais, porém não-conceituais. Tais estados podem acompanhar outros estados intencionais de conteúdo proposicional. Podem servir de *background* para processos cognitivos envolvendo estados intencionais de conteúdo proposicional. O decisivo, porém, é que os estados fenomenais intencionais *não* são eles mesmos conceituais. Sensações corporais como, por exemplo, a dor de dente ou o orgasmo são representacionais no sentido em que refletem causalmente, *em circunstâncias normais*, as alterações em partes específicas do corpo. Estas alterações formam o conteúdo representacional das sensações corporais. Tais representações (i) são o *output* de receptores nervosos respondendo a estímulos diversos no nosso corpo e (ii) integram o *input* dos processos cognitivos em que são, pela primeira vez, reveladas cognitivamente ao sujeito das representações. Tye diz, nestes casos, que mediante a ação de processos cognitivos operando sobre as representações não-conceituais o sujeito se torna consciente de (*aware of*) seus estados fenomenais. Tornar-se consciente de (*to become aware of*) suas próprias experiências é um processo cognitivo que subsume, pela primeira vez, a experiência fenomenal a conceitos. Só então o sujeito experiencial adota uma atitude epistêmica (cognitiva) frente a seus estados fenomenais. *Abaixo* deste nível de *awareness*—chamado tradicionalmente de “introspecção”—está a consciência fenomenal (*consciousness*). O conteúdo fenomenal seria, portanto, mais bem descrito como representações não-conceituais de alterações físicas prontas a integrar o *input* intencional de processos que têm como *output* outros estados intencionais de conteúdo proposicional (*beliefs* e *desires*). Ser o sujeito de uma experiência fenomenal significaria assim ter, no *input* de certos processos cognitivos, representações adicionais, representações não-conceituais que poderiam promover respostas diferenciadas.

A explicação fornecida pela teoria PANIC nos permite acomodar o fisicismo à intuição de que a experiência de *ver vermelho* e de *saber o que é ver vermelho* sejam experiências distintas que compartilhem um mesmo conteúdo proposicional. Ou seja, algo de novo parece ocorrer a algum que *vê* pela primeira vez a cor vermelha, ainda que esta mesma pessoa já soubesse tudo o que fosse possível saber, em termos fisicistas, sobre o que seja *ver vermelho*. Aqui esbarramos no núcleo de um dos principais argumentos contra o fisicismo, o argumento do conhecimento. Para concluir, pretendo mostrar como a descrição da consciência fenomenal proposta pela teoria PANIC permite responder, sem qualquer restrição ao fisicismo, ao que teria ocorrido de *novo* na experiência de Mary.

Mary é a famosa cientista de Frank Jackson que conhece tudo sobre o mundo físico. Mary possui, por assim dizer, a totalidade do saber acerca dos elementos e leis da física do futuro através da qual definimos o próprio fisicismo. O conhecimento de Mary compreende, portanto, tudo o que pode ser compreendido no vocabulário fisicista. Após a apresentação da nossa personagem, vamos à estória. A brilhante Mary cresceu e apreendeu tudo o que sabe em um quarto preto-e-branco. Todo o seu contato com o mundo exterior se dá através de um monitor preto-e-branco e, para contribuir com o argumento, podemos imaginar que algo misterioso no quarto faz com que a própria Mary se veja em preto-e-branco. Um dia Mary é retirada do quarto e vê, pela primeira vez, tomates vermelhos, ou seja, vê, pela primeira vez cores. Mary já sabia que não apenas tudo a respeito de tomates, como também acerca da experiência de cores. Já sabia, portanto, que tomates maduros, hidrantes e cabines telefônicas inglesas são vermelhos. Já sabia também tudo acerca dos processos fisiológicos que acompanham a visão de cores. Neste ponto o fisicismo é então posto a prova: Supondo que compartilhemos a intuição de que algo de novo ocorreu quanto Mary viu cores pela primeira vez, não deveríamos também admitir haver algo na nossa experiência, mais particularmente, na nossa vivência de certas experiências que não pode ser descrito em termos fisicistas? O argumento de Mary parece apontar exatamente para a suposta distinção entre conhecer e vivenciar algo. Mary conhece tudo sobre as cores, mas nunca, antes de deixar o quarto, teve ela mesma a experiência de ver cores. Neste sentido, embora saiba o que é ver cores, nunca soube o que é ser alguém que vê cores. O que parece irreduzível ao fisicismo é, portanto, mais uma vez este caráter privilegiado da perspectiva de quem é o sujeito da experiência, em outras palavras, o caráter subjetivo da experiência.

Acompanhando o argumento, pretendo aceitar os diversos elementos que compõem o caso, por mais fantásticos que estes possam parecer. Não pretendo também negar a intuição básica de que Mary viveu algo novo ao sair do quarto. O que sim pretendo é negar que este algo novo não possa ser descrito em termos fisicistas, ou que represente o limite do próprio fisicismo. Se Mary já sabia tudo sobre tomates vermelhos obviamente o que ocorreu com Mary ao vê-los pela primeira vez em nada pode alterar seu conhecimento. Porém podemos dizer que algo de novo é introduzido na geração do conhecimento em questão. De acordo com a teoria proposta por Tye, novas representações, não-conceituais, passam a integrar o *input* dos processos cognitivos de Mary. Embora nada alterem do conteúdo cognitivo de Mary, tais representações poderão promover alterações no seu comportamento futuro. Poderão, por exemplo, habilitá-la a selecionar com maior rapidez e eficiência os tomates que deseja colocar na salada. Os elementos introduzidos apontam para fatos que Mary já conhecia, posto conhecer tudo acerca do mundo físico. Porém sua introdução permite uma reorganização dos dados informacionais de que Mary dispõe e que determinam o seu comportamento. Segundo tal interpretação, o que Mary adquire não é um fato novo, mas uma nova forma de acesso a fatos por ela já conhecidos. Assim sendo, não haveria nem fatos que não pudessem ser cobertos por uma teoria fisicista do mundo,

nem acesso não-físico aos mesmos. A hipótese fisicista sairia, mais uma vez incólume.

Se a descrição funcional de nossos estados mentais, em geral, e do suposto caráter subjetivo da experiência, em particular, for plausível, não estaríamos eliminando, não apenas o gap explanatório, mas também o vocabulário mentalista e práticas de explicação até então aceitas e bem sucedidas? Este é o ponto em que me parece razoável admitirmos estar diante de um dilema. Talvez este seja também o ponto que separa uma perspectiva meramente eliminativista de uma explicação funcional do mental. Não queremos simplesmente suprimir o mental, mas explicá-lo de acordo com as regras do fisicismo. Mas se conseguirmos oferecer uma explicação convincente - que no futuro possa se comprovar correta -, parece que já não sobrará espaço para uma explicação de outro tipo. Nossas práticas explicativas parecem assim entrar em um progressivo processo de substituição. A causa do comportamento desordenado de um epilético já não poderia comportar ao lado de si outra causa que não a própria epilepsia como síntese de uma série de processos neurofisiológicos. Possessão demoníaca ou complexos familiares são alternativas banidas pela aparente descobertas das *verdadeiras* causas do comportamento epilético.

Neste ponto, alguns de nossos mais respeitáveis interlocutores parecem pular.² Até aqui o fisicismo poderia parecer não representar uma autêntica ameaça a práticas hoje bastante consagradas. Mas quando as práticas com que estamos familiarizados se veem ameaçadas, o feitiço parece virar contra o feiticeiro: que tipo de perspectiva é esta que - sem que possa fornecer uma explicação mais substancial do mental, apelando apenas a possibilidades explicativas futuras - supõe poder colocar em xeque nossas práticas cotidianas bem sucedidas e consagradas ao longo do tempo? Aqui, nossa única resposta deve ser: trata-se de um fisicismo *prima facie*, uma aposta na futura compreensão funcional do mental que venha a preencher lacunas explicativas atuais e compatibilizar nossa descrição de projeto humano com nossas convicções materialistas.

Notas:

* Este artigo é uma homenagem a Nicolas Kaufmann.

¹ Sobre a defesa do compatibilismo com base em uma exposição do modelo do autodomínio, ver: Simon Blackburn (1999).

² Penso aqui em autores como Lynne Rudder Baker que, embora assumidamente materialistas, procuram resgatar a credibilidade de práticas explicativas cotidianas envolvendo fatos psicológicos, econômicos e sociais. Sobre este tema ver Baker (1995).

Tying Phenomenal Consciousness and Quining the Gap⁵²*

In his article about the problem of consciousness and the explanatory gap⁵³, J. N. Kaufmann presents interesting contributions to the discussion of the qualitative characteristics of subjective (or first-person) experiences of mental phenomena. Basically, the problem emerges from the fact that; in one hand, those characteristics

* I would like to express my thanks to Carlos Eduardo Nogueira Loddo (UQTR) for his extremely useful comments on earlier versions of this paper. His stimulating observations helped me to present my points in a clearer way, especially as far as conceptual matters are concerned. He also helped essentially in formulating a precise interpretation (in what is relevant for this discussion) of the perspective proposed by J. N. Kaufmann, with whom he worked closely, translating many of his articles (including the one here commented) to Portuguese. For the English revision of the present text, I would like to harmly thank

⁵² Of course, the use of the neologism "to quine" is inspired in Dennett's use of it, but in my usage what is being eliminated is not an ontological element, but rather the explanatory gap, which is an element of an epistemological and conceptual problem. In terms of ontology, the entities that supposedly generate the problem (the qualitative characters of subjective experience and the correspondent phenomenal consciousness) are retained in the approach proposed. The basic elements of this approach are proposed by Michael Tye, therefore I am using another new term, "to tye", taking advantage of the phonetic coincidence between Tye's name and the English verb, "to tie". There is also a semantical coincidence. Tye's approach permits phenomenal consciousness to be kept in the ontology, therefore preserving its causal role, without confusing it with cognitive contents. It is somewhat like "to tie" phenomenal consciousness to the other elements of the ontology, and "to tie" its concept into explanatory structures, by the same token.

For Dennett's use of "quining", see Dennett, D. (1998) "Quining Qualia", in *Consciousness in Contemporary Science*, eds. A. Marcel and E. Bisiach (Oxford: Oxford University Press, 1988), 43-77. Cf., also: *The Philosophical Lexicon*, ed. D. Dennett. Blackwell Publishers, 2002. The term is evidently inspired in the American Philosopher W. V. O. Quine. According to the *Lexicon's* definition, 'to quine' means "To deny resolutely the existence or importance of something real or significant".

⁵³ Kaufmann, J.N., "A problemática da Consciência e a Lacuna Explicativa", in *Princípios. Revista de Filosofia*, Universidade Federal do Rio Grande do Norte, Brazil, v.6, n.7, pp.87-120, 1999.

seem to play a causal role in human actions and, in the other hand, no conceptually satisfactory reference to those elements seem to fit (due to their subjective character) into explanatory structures (due to the demands of objectivity associated to these structures). Therefore, there seems to be a gap between the qualities of subjective experience and acceptable psychological explanation of human action, in a physicalist framework. Kaufmann accepts the formulation of the problem in J. Levine's terms.⁵⁴

If such a gap would be the case, it would consist in a serious problem to any naturalization project of psychological explanation. The problem would be a serious one, because it is not simply a matter of fact, relative to the present state of our science (which could change in the future), but a *matter of principle*. Being the explanatory gap problem formulated as a matter of principle, Kaufmann observes that it should be also seen as a *conceptual* matter.

Because of this *conceptual* character of the problem, Kaufmann starts out his analysis by attempting to circumscribe precisely the *concept* of consciousness about which the problem is. He first presents a series of concepts and conceptual oppositions concerning consciousness, and establishes which concept of consciousness is the relevant one for this discussion. Especially important for this, is N. Block's opposition between *access consciousness* and *phenomenal consciousness*.⁵⁵ According to such opposition, *access consciousness* is representational and definable functionally. *Phenomenal consciousness*, according to this account, cannot be defined as a general concept, although it is possible to indicate which objects would fall under it. Kaufmann observes that the list of such objects vary, from author to author. But this last concept is, anyway, the notion of consciousness that is involved in first-person or subjective experience. Therefore, Kaufmann locates here the concept of consciousness around which the problem of the explanatory gap is formulated, and relatively to which it should be treated.

Next, Kaufmann goes on to examine two contrasting perspectives, relatively to the explanatory gap problem. First, he examines C. McGinn's position, which includes arguments favorable to the idea that there is a gap.⁵⁶ Kaufmann examines two of McGinn's arguments for the gap, using O. Flanagan's reconstructions of those arguments.⁵⁷ Kaufmann retains Flanagan's criticisms to the arguments and adds to those criticisms some others, taken from E. Husserl's phenomenology.

⁵⁴ J. Levine, "Materialism and Qualia: the explanatory gap". *Pacific Philosophical Quarterly*, 64, 1983.

⁵⁵ Block, N. "On a Confusion About a Function of Consciousness." *Behavior and Brain Science* 18(2): 227-247; 272-289, 1995.

⁵⁶ C. McGinn, *The Problem of Consciousness*. Oxford: Basil Blackwell, 1983 and 1991

⁵⁷ O. Flanagan, *Consciousness Reconsidered*. Cambridge, M.A.: MIT Press, 1992.

However, Kaufmann's exam of D. Chalmers's position is the relevant one to what we will discuss in this text. Chalmers's position is in favor of "closing the gap", this is to say, he tries to provide a solution to the problem. As Kaufmann observes, Chalmers presents different strategies for this, in different texts. In the early texts⁵⁸, Chalmers attempts to supplement physicalist explanations with conceptual analysis typical from phenomenology.

In a later text⁵⁹, in view of correlating consciousness to neuronal processes (this correlation being the conclusion of an argument he formulates), Chalmers presents what he calls "bridging principles". Functionally definable states are linked to neuronal structures, because the first might be realized in the last (the empirical premise of Chalmers's argument). We might have access to consciousness by means of reports or by other behavioral modes of expression (here is the bridging principle of Chalmers's argument). Kaufmann analyses Chalmers's argument, presenting a series of criticisms to it. In some cases, Kaufmann takes into consideration Husserl's phenomenology, in order to formulate those criticisms.

It is not my purpose here, neither to examine Chalmers's argument, nor to provide a full account of Kaufmann's criticism to it. However, I will retain a single and crucial point observed by Kaufmann. He notices that Chalmers's bridging principles may only have success in linking physicalist explanations to consciousness, *if* consciousness is conceptualized as what Block calls *access-consciousness* (the one which is representational and definable functionally), but *not at all* if it is conceptualized as *phenomenal-consciousness*.

This last concept, as seen, is where the explanatory gap problem lies, according to Kaufmann's approach. Even though this observation is a critical remark towards Chalmers's attempt to close the gap, Kaufmann's attitude is more optimistic than that, as far as the project of closing the gap is considered. Kaufmann observes that, *if* Chalmers would show (a possibility we do not find in Block's account) that a concept of functionally defined consciousness is *systematically correlated* with a concept of phenomenal consciousness, *then* his argument could perhaps stand, and the gap closed. I would endorse Kaufmann's main framework.

In the following lines, I will examine a perspective that seems to be able to close the gap. By following it, I will try to argue in the sense that there is no gap

⁵⁸ Chalmers, D. J. *The Conscious Mind : In Search of a Fundamental Theory*. New York, Oxford University Press, 1996 ; Chalmers, D. J. "Facing up to the Problem of Consciousness" – "Response". *Explaining Consciousness--the Hard Problem*. J. Shear. Cambridge, MIT Pr: 9-33; 379-422, 1997.

⁵⁹ Chalmers, D. J. "On the Search for the Neural Correlate of Consciousness". *Toward a Science of Consciousness II : The Second Tucson Discussions and Debates. Complex Adaptive Systems*. S. R. Hameroff, A. W. Kaszniak and A. Scott. Cambridge, Mass. ; London, MIT Press: xx, 764, 1998.

between physicalist explanations and phenomenal consciousness, as a matter of principle. This perspective is the one suggested by Michael Tye.⁶⁰ In a sense, what Tye proposes corresponds to Kaufmann's suggestion to save Chalmers's argument, since Tye does correlate systematically a concept of phenomenal consciousness with a concept of functionally definable consciousness. According to Tye, more radically, the concept of phenomenal consciousness is functionally definable, **although it is not definable in the same way as conceptual contents. Anyway, those elements are systematically related. Tye's trick is to consider that phenomenal consciousness (differently from what we find in Block's account) is functionally definable, although (like in Block's account) it is not representational.**

Based on an analysis of phenomenal consciousness and of certain *sui generis* properties in these lines, I intend to take a stand, relatively to the *explanandum* of the problem in a way that the apparent explanatory gap between the mental and the physical is eliminated. More specifically, I want to defend the thesis according to which the so-called *sui generis* qualities of mental events, referred to as *qualia*, point to physical events with peculiar characteristics. These events would constitute, as Tye puts it, the nonconceptual representational contents of the input of physical processes whose output would be other intentional states with propositional content which are susceptible, at least *prima facie*, to take part in a functionalist explanation.

We can understand physicalism, in general, as the perspective according to which everything that exists or that may be known belongs to the physical world, i.e., in a strict sense, belonging to the domain of objects and laws of an elementary microphysics, possibly to be developed in the future.

Whichever definition of the “physical world” we would adopt, it would seem that it should incorporate a set of properties applicable in fact to a specific group of objects that, at least on a first impression, cannot be imputed meaningfully to physical objects. In this sense, it seems that I would be the only possible subject of my pain, which would exist only when it is felt by me.

For instance, we say that “no one else can feel my pain” as well as we believe that pain, sadness, gladness and other mental states just exist while lived by a subject or while belonging to a subject. My consciously perceptible experience, my body sensations, emotions, passions and moods seem to be reported to me in a way that my car and my clothes cannot be. Another person could have my car and my clothes and those items would not cease to exist if not possessed by someone.

The “monopolization” of experienced mental phenomena by a subject, by a self, seems to characterize a sort of inner belonging, which is alien to the objects that only contingently belong to me. Should we not distinguish the objects to which I have a privileged access, that is, the objects that can only be “possessed” and “known” in

⁶⁰ Michael Tye, *Ten Problems of Consciousness*. Cambridge, Mass.: MIT Press, 1995.

my experienced perspective and that cannot exist, unless they are related to the subject of the experience – from the other objects to which I simply relate to, because I am the one that contingently possesses them? If the characteristic properties of the first group - that is, the property of privacy, the property of being necessarily owned and the property of perspectivity⁶¹ - cannot be attributed to physical objects, should we not then admit the existence of nonphysical entities?

There would be problems, if we would adopt such an assumption. With such an assumption; and keeping together the supposition that the physical world is causally closed, i.e., all causes of physical events being physical events; we would be unable to establish causal relations between our sensations, emotions, passions and moods and physical events. At best, phenomenal events could be effects but never causes of physical events. But are we really prepared to admit that phenomenal states are causally irrelevant?

The apparently unavoidable way in which we react to the presence of these properties seems to establish a gap between the physicalist explanation and what we innerly believe to happen in some contexts. Would it be possible to explain in a physicalist vocabulary, that is to say, in terms of elementary microphysics, even if in the future, what it means to be the subject of some experiences, or to express somehow the privileged point of view that we assume towards our sensations, emotions, passions and moods?

Facing this problem, we seem to find three alternatives:

1. to accept the *sui generis* character of phenomenal objects and their irreducibility to the domain of the physical world;
2. to show that phenomenal objects do not really have these *sui generis* properties; or to show that these objects do not exist and, thus, that we cannot apply these properties; or
3. to acknowledge the existence of phenomenal objects and their characteristic properties denying, however, that they do not apply to physical objects;

The first alternative is clearly dualist, that is, it reinstates an ontological dualism in which the relation between the mental and the physical becomes contingent, arbitrary or even impossible. At best, it would be possible to restore an epiphenomenalism, according to which the mental would be causally inert. Our 21st century materialist beliefs, however, makes it difficult to understand what objects that do not produce effects on the physical world would be. In other words, to claim the

⁶¹ About these properties, see: Michael Tye, (1995).

existence of something seems to imply causal relations with other physical objects or events.

The second alternative is either reductionist or eliminativist. Respectively, it either reduces the phenomenal objects to physical ones, thereby denying their peculiar characteristics; or it denies the existence of these objects and consequently the realization of *sui generis* properties. This alternative leaves physicalism untouched, but not our ordinary practices of attributing causal powers to mental states.

The third alternative seems to combine physicalism with some kind of dualism or pluralism of properties, thus affording some understanding of the peculiar character of some of our experiences. The second and the third alternative share the kernel of the physicalist intuition. The difference between them is subtler than it seems to be at first sight. Some sort of reductionism or of eliminativism seems unavoidable even to the third alternative. The difference between them lies only in the attempt of the third alternative to conciliate physicalism with some of our most valuable intuitions about our subjective states. But which subjective states, specifically? Or, in other words, which of their peculiar characteristics can plausibly be combined with the kernel of our physicalist beliefs?

Here we need to clarify our position in relation to the *explanandum* of the problem. What seem to resist to a physicalist reduction are not mental states in general, but some of their supposed properties. In particular, these are the properties of: privacy, necessary ownership and perspectival subjectivity. These properties reflect the subjective character of our experience, that is, they are connected to what is lived only and exclusively from the perspective of the experienced subject. To use Nagel's phrase, those properties constitute the "what it is like to be"⁶² character of our experience. In Nagel's view, these properties are present in every intentional event, i.e., in every kind of relation that would involve consciousness and its objects: they constitute the irreducible subjective character of intentional events.

Here I sustain a more restricted position: a position that excludes from this domain our mental states such as beliefs, desires and intentions whose relevant content is a propositional attitude. In order to justify such a restriction, I intend to show that it is possible to provide, at least *prima facie*, a functional description of our propositional attitudes and that such a description seems not to leave out any relevant element of our comprehension of these mental states. The irreducible subjective character of experience, or *quale*, will be seen as a peculiar quality of some mental states. Thanks to these peculiar qualities, the perspective of the experiencing subject adds something irreducible to the experience itself. I would, therefore, like to

⁶² Thomas Nagel, "What it is like to be a bat", in: *Mortal Questions*. Cambridge: Cambridge University Press, 1979.

characterize the problem of the explanatory gap as one directed specifically to mental events that seem not to be sufficiently described without a direct reference to the agent.

Before we reach the problem of the gap itself, I shall give a brief exposition of what could be a functional explanation of our propositional attitudes thus justifying the restriction mentioned above.

The functionalist perspective in general identifies mental states with the function accomplished by brain states or with brain states that accomplish that function. What permits a brain state to be associated to a content p (e.g., the content of a belief) is its relation with certain sensorial stimuli, with other inner states and with behavior. Therefore, mental states are both inner causes of behavior and effects produced by the external world. In the words of Jackson and Braddon-Mitchell:

*"... A functionalist theory of mind specifies mental states in terms of three kinds of clauses: input clauses which say which conditions typically give rise to which mental states; output clauses, which say which mental states typical give rise to which behavioral responses; and interaction clauses, which say how mental states typically interact."*⁶³

So as not to beg the question, the functionalist characterization of each state must be in exclusively in physical terms, that is, the intentional states that we intend to describe in physical terms must not be reintroduced in the characterization of the output. Outputs must be described entirely in physical terms. The functional description of mental states specifies a net of physical relations, involving dispositions to behavior, typical causes, as well as other mental states as far as these are also characterized in functional terms.

The most obvious applications of the theory follow the model of a machine programmed to perform specific tasks. This includes very simple machines like the ones that provide us with Coke when we introduce a coin of a certain value, but also highly complex ones like connectionist machines of the PDP kind or neural nets. Evidently, the complexity of the functions proposed by mental states requires flexible and very complex models. In many cases the complexity is so as that we are as yet unable to specify the models in question. So, the functionalist proposal seems vulnerable to an objection in fact, but if we are right, it may show that, in principle, there is nothing to prevent relevant aspects of mental events from being described in physical terms or in physically realizable terms.

The human mind described in a functional way thus operates like a computing machine composed of several modules. There would be a scanner in the first module,

⁶³ David Braddon-Mitchell and Frank Jackson, *The Philosophy of Mind and Cognition*, 1996, p.40.

which is responsible for the reception of the inputs. From that point on, we can imagine several modules. Among them, there would be an evaluator module, responsible for the selection of information that will get to the final stage, that is, the performance of a specific behavior. The peculiarity of such a program is its capacity to learn, that is, to alter the final result, due to new inputs that would involve the effects produced by the agent behavior in the previous stages. In other words, the produced output would promote external answers, which, in their turn, would be introduced in the input and evaluated by the program, so that it can provide a new result. Thus, e.g., a program able to recognize alphabetic characters, like the ones successfully used in postal services, transforms and improves its performance when different graphics of the same character are provided and the manager answers affirmatively or negatively to its recognition. Such programs are capable of organizing the information so that character previously ignored can later be recognized. Based on the functional analysis of software program models, we could therefore provide an account for one of the main characteristics of the human model: our capacity to learn. Thus, other typically human characteristics, such as our capacity to deliberate and to intervene in the course of our actions – on which depends our attribution of responsibility and freedom to an agent – could also be recovered successfully.⁶⁴

Supposing, this way, that a functional description of our propositional attitudes is possible and with such explanation nothing essential to them is lost, the problem of the explanatory gap, or the problem of the physicalist reduction, becomes limited to the question of *qualia*. How do we fit them in a physical description of the world? What exactly do they refer to?

If we are not ready to accept a reductionist or an eliminativist perspective that refuses in advance that something beyond linguistic conventions may ground out beliefs in such properties and in the objects to which they supposedly apply, we must be able to provide a physical our physically realizable description of them. Or, in other words, to explain in physicalist terms how “to be the subject of certain experiences” (what it is like to be) may play some functional role in our behavioral dispositions. This possibility is what I intend to analyze based on the theory proposed by Michael Tye: “I call this the PANIC theory of phenomenal character: phenomenal character is one and the same as Poised Abstract Nonconceptual Intentional Content” (137).⁶⁵ Even though Tye’s PANIC theory is constituted of two aspects (a **nonconceptual** one and an **abstract** one), I will, from this point on, consider exclusively the **nonconceptual** aspect of it, which seems to me the relevant one for the present discussion.

⁶⁴ About a defense of compatibilism based on the exposition of the model of self-dominion, see: Simon Blackburn, *Think*, Oxford University Press, 1999, p.98.

⁶⁵ Tye, M. (1995) *Ten Problems of Consciousness*, p.137.

The PANIC theory is a theory of phenomenal events or states as intentional but nonconceptual phenomena. These states may come together with other intentional states with propositional content. They can work as background elements to cognitive processes involving intentional states with propositional content. The crucial point here is that the intentional phenomenal states are not themselves conceptual.

Body sensations, such as a toothache or an orgasm are representations, in such a way that they reflect causally, in normal circumstances, the alterations in specific body parts. These representations (1) are related to the output of nervous receptors responding to several stimuli in our body and (2) integrate the input of the cognitive processes that are for the first time cognitively revealed to the experience subject.

In these cases, Tye says that through the action of cognitive processes operating in the nonconceptual representations the subject becomes aware of his phenomenal states. Becoming aware of his own experiences is a cognitive process that submits for the first time the phenomenal experience to concepts. Only at this moment, the experiential subject does adopt an epistemic (cognitive) attitude in relation to his phenomenal states. Within this level of awareness – traditionally called introspection – there is the phenomenal consciousness. The phenomenal content would therefore be better described as a nonconceptual representation of physical alterations, ready to integrate the intentional input of processes that have as output other intentional states with propositional contents (beliefs and desires). This way, to be the subject of a phenomenal experience would mean to possess certain nonconceptual intentional contents, which might be inputs of certain cognitive processes, that is, nonconceptual representations that could provide different responses.

The explanation provided by the PANIC theory allows us to make physicalism compatible with the intuition that the experience of *seeing red* and the experience of *knowing what it is to see red* are distinct experiences that share the same propositional content. In other words, something new seems to occur to someone that sees red for the first time even if this very same person already knows everything that is possible to be known, in physicalist terms, about *what it is to see red*. Here we face the case of one of the main arguments against physicalism, the knowledge argument. In what follows, I should like to show how the description of the phenomenal consciousness proposed by the PANIC theory allows us to account for Mary's new experience without jeopardizing physicalism.

Mary is Frank Jackson's famous scientist who knows everything about the physical world.⁶⁶ She has complete knowledge of the elements and the laws of future's physics, by which we have previously defined physicalism. Mary's knowledge involves therefore everything that is possible to be known in the physical

⁶⁶ Jackson, F. (1986). "What Mary Didn't Know." *Journal of Philosophy*: 83 291-295.

domain. After having introduced Mary, let us take a look at her story. Bright Mary has grown up and learned everything in a black-and-white room. All of her contacts with the external world have been through a black-and-white monitor. For the sake of the argument, we can imagine that something mysterious in the room makes her see herself in black-and-white. One day Mary leaves the room and she sees for the first time red tomatoes, that is to say, she sees colors for the first time. Mary already knew not only everything about tomatoes but also about color experience. She knew, therefore, that ripe tomatoes, hydrants and British telephone booths are all red. She also knew everything about the physiological processes that follow color experience. At this point, physicalism is challenged: supposing that we share the intuition that Mary learned something new when she saw colors for the first time, should we not admit that there is something in our experience that cannot be described in physicalist terms? Mary's argument seems to indicate exactly the distinction between *to know something* and *to experience something*. Mary knows everything about colors, but she never had any experience of seeing colors, before leaving the room. This way, although she knows *what it is to see colors*, she has never known *what it is like to see colors*. What seems to be irreducible to physicalism is therefore this privileged character of the perspective of the person who is the subject of experience, in other words, the subjective character of the experience.

For the sake of the argument, I shall accept the elements of the story, although they may seem very implausible. I do not intend to deny the basic intuition that Mary experienced something new when she left the room. What I do want to deny is that this new something cannot be described in physicalistic terms or that it indicates the boundaries of physicalism. If Mary already knew everything about red tomatoes, obviously what happened to her when she first saw them could not have altered her knowledge. But we can say that something new is introduced in the generation of such knowledge. According to Tye's theory, new nonconceptual representations come to integrate the input of Mary's cognitive processes. Although they do not alter Mary's cognitive content, such representations can promote changes in her future behavior. They can, for instance, provide her with the skill to select more quickly the tomatoes she wants to put in a salad. The introduced elements indicate facts that Mary already knew, since she knew everything about the physical world. But, the introduction of these new elements, allow a reorganization of the data available to Mary that determines her behavior. According to such an interpretation, what Mary acquires is not a new cognitive content, but a new form of access to cognitive content she already possessed. So, there would not be cognitive contents that could not be covered by a physicalist theory of the world, neither nonphysical access to them. The physicalist hypothesis would be kept untouched.

If the functional description of our mental states *in general*, and the above description of the supposedly subjective character of our experience *specifically*, are both plausible, would we then not only eliminate the explanatory gap but also mentalist vocabulary and some widely accepted and successful explanatory practices?

This is the point where it seems to me reasonable to admit that we are facing a dilemma. Maybe it is also the point that separates a mere eliminativist perspective from a functional explanation of mental states. We do not simply want to suppress the *sui generis* characteristics of our experience but we want to explain them from a physical point of view.

But if we are successfully offering a plausible explanation – which can be proven correct in the future – it may leave no room for other sorts of explanation. Our explanatory practices seem to enter in a progressive process of substitution. The cause we are ready to admit as attributed to the disorganized behavior of an epileptical patient could no longer be admitted, together with another one. We might admit epilepsy itself in a causal story, conceptualized functionally, and associated by realization links to neurological processes. Devil possession or family complexes, in this case, are suppressed alternatives. They are not acceptable hypothesis for the *true* causes of the epileptical behavior.

Some philosophers seem to skip this point.⁶⁷ Up to now, physicalism could hardly threaten very well-established explanatory practices. But when practices we are familiar with are threatened, the question returns against us: What sort of perspective is the one supposed to challenge our day by day successful practices without providing a substantial explanation of mental events and claiming only future explanatory possibilities? Here our only answer should be: it is a *prima facie* physicalism, a bet in the future functional comprehension of mental events that will fill actual explanatory blanks and will make our description of the human design compatible with our materialist beliefs.

⁶⁷ I think here about authors like Lynne Rudder Baker who, although an assumed materialist, intends to rescue the credibility of day by day explanatory practices which have in their causal antecedent psychological, economical and social facts or events. See L. R. Baker, *Explaining Attitudes*. Cambridge: Cambridge University Press, 1995.

Quem somos nós? Pressupostos e consequências do programa de naturalização do Self

Da antiguidade até os dias de hoje a filosofia tem se empenhado exaustivamente em esclarecer a atuação do homem no mundo. Seu papel tanto como sujeito e produtor de conhecimento, como agente moral e transformador da realidade, tem sido sempre, de forma mais ou menos direta, relacionado ao fenômeno da consciência. A constatação de um acesso privilegiado aos nossos próprios estados mentais, fez com que nossa identificação enquanto *res cogitans* fosse assegurada como a única certeza num mundo manipulado por um possível “gênio maligno”. Com o cogito é resgatada também a veracidade de todos os seus modos de existência. Somos assim identificados, antes de tudo, com o conjunto de nossos estados mentais: o pensar, o duvidar, o desejar etc. Apenas depois de um longo e tenebroso império da dúvida hiperbólica, resgatamos, também, nossa identidade como *res extensa*. De apenas uma mente, passamos a ser então uma mente e um corpo e com isto instauramos um problema que até hoje ocupa uma posição de destaque na filosofia: qual a relação entre estas duas partes, mente e corpo, com as quais, aparentemente, estamos identificados? Uma resposta satisfatória a esta questão deveria não apenas esclarecer o que somos, como justificar o *status* especial que temos reclamado para nós, humanos, no mundo ao qual pertencemos.

No âmbito das discussões morais a pergunta metafísica pelo que somos, embora quase sempre calada, parece determinar nossas inclinações teóricas, nosso endosso ou recusa de práticas específicas. Seremos ou não os filhos de Deus, termos ou não uma alma imaterial e imortal, uma consciência transcendental e uma vontade livre de todas as determinações materiais parece determinar não apenas o que julgamos razoável exigir dos seres humanos, mas, sobretudo, para os seres humanos. É neste sentido que nos reportamos inúmeras vezes à chamada “dignidade humana” para avaliar o valor moral de práticas como aborto, eutanásia, clonagem etc. Mas o que seria tal dignidade? Com certeza, aquilo que nos torna especificamente, humanos. Com isto retornamos à questão inicial. O que realmente somos? Por que ou como podemos justificar o tratamento diferenciado que reclamamos perante todas as demais entidades?

Nesse artigo busco conciliar a adoção de uma perspectiva moral específica com convicções naturalistas. Para tal, em primeiro lugar apresento um modelo do que seria uma concepção naturalista do ser humano e, em seguida, verifico até que ponto tal concepção é ou não capaz de resgatar de forma abrangente e coerente nossas diversas convicções morais.

1. A naturalização da mente

Nesta etapa, pretendo defender uma perspectiva funcionalista da mente e, por conseguinte, uma concepção materialista do ser humano. Dentro desta perspectiva, defenderei que aquilo a que atribuímos às propriedades mentais não pode ser identificado nem como uma entidade não-física, nem com o cérebro. Com isso pretendo suprimir o dualismo mente-corpo, tanto em sua versão cartesiana, como na versão materialista contemporânea. Propriedades mentais, assim como as demais propriedades físicas, são atribuídas à sistemas e às funções realizadas pelos mesmos.

Para defender uma perspectiva funcional da mente, pretendo, em primeiro lugar, analisar nossas intuições acerca do caráter peculiar dos eventos mentais e, em seguida, inseri-los em uma descrição estritamente funcional.

Por naturalização da mente podemos compreender, em linhas gerais, a perspectiva segundo a qual tudo o que existe ou que pode ser conhecido pertence ao mundo natural: ao domínio constituído pelos objetos e leis descritos pelas ciências do mundo físico. Qualquer que seja a definição adotada de “mundo físico” parece, no entanto, haver um conjunto de propriedades *sui generis* aplicáveis a um grupo específico de objetos que, pelo menos à primeira vista, *não podem* ser atribuídas significativamente aos objetos físicos. Aceitamos como uma descrição fenomenológica adequada do caráter inalienável de nossos próprios estados ou de nossa relação para com nós mesmos algo assim como é expresso nos versos de Caetano Veloso: “Cada um sabe a dor e a delícia que é ser o que é”. Dizemos que “ninguém mais pode sentir a minha dor,” assim como também acreditamos que dor, tristeza, alegria e demais estados de espírito só existem enquanto são vivenciados ou enquanto pertencem a um sujeito. É neste sentido que pareço ser o sujeito único da minha dor e que a mesma pareça só existir quando referida a mim. Minhas experiências perceptivas conscientes, minhas sensações corporais, emoções, paixões e humores parecem reportar-se a mim de tal forma que meus demais pertences - meus livros, minhas fotos, minha casa e minhas roupas, por mais personalizados e relacionados à minha personalidade que eu queira interpretá-los - efetivamente não o fazem. Outra pessoa poderia vir a possuir a minha casa ou se apossar de meus livros, mas estes não deixam de existir tão logo deixem de ser posse de alguém. A “monopolização” de fenômenos mentais por um sujeito, por um “eu,” parece caracterizar uma forma de pertencimento intrínseco alheio aos demais objetos físicos que estão, contingentemente, sob o meu poder. Não deveríamos, então, distinguir os objetos aos quais tenho um acesso privilegiado, enquanto sou aquele que os *vivencia intrinsecamente*—isto é, os objetos que só podem ser “tidos” e conhecidos na *minha* perspectiva experiencial e que não subsistem a não ser em relação com sujeitos de experiência—de outros objetos aos quais me relaciono simplesmente porque sou aquele que os *tem* contingentemente? Como compatibilizar a premissa

naturalista/fisicista⁶⁸ com nossa crença cotidiana de que são estes estados peculiarmente atribuídos a um agente, e não meras leis do mundo físico, o que determina o que somos e nossa atuação no mundo?

O modo aparentemente incontornável com que somos tomados por tais convicções e que reagimos a aparente evidência de tais propriedades parece instaurar um fosso entre as explicações naturalistas/fisicistas e o que mais intimamente supomos ocorrer em certos contextos. Seria possível explicar no vocabulário fisicista o que é ser o sujeito de determinadas vivências ou retratar este ponto de vista privilegiado que inevitavelmente assumimos diante de nossas sensações, emoções e humores?

Diante deste problema, restam-nos três alternativas:

1. Aceitar o caráter *sui generis* dos objetos fenomenais e sua irreducibilidade ao domínio do mundo físico;
2. Mostrar que os objetos fenomenais *não* possuem realmente as propriedades *sui generis* mencionadas; ou mostrar que os mesmos não existem e, por conseguinte, que tais propriedades não se aplicam;
3. Reconhecer a existência de objetos fenomenais e de suas propriedades características, negando, porém, que estas não se apliquem também aos objetos físicos.

A primeira alternativa é claramente dualista, ou melhor, reabilita um dualismo ontológico no qual a relação entre o mental e o físico torna-se contingente, arbitrária ou mesmo impossível. Sob a suposição de que o mundo físico é causalmente fechado - ou seja, que para todo evento físico deva haver ao menos uma causa física - estaríamos então impossibilitados de estabelecer relações causais entre o grupo de eventos aos quais supomos pertencer nossas sensações, emoções, paixões e humores, por um lado, e os eventos físicos, por outro. Na melhor das hipóteses, os eventos mentais poderiam ser efeitos, mas jamais causas, de eventos físicos, ou seja, poderíamos assumir uma perspectiva epifenomenalista, mediante a qual o mental seria considerado causalmente inerte. Mas estaremos realmente dispostos a admitir que estados fenomenais sejam causalmente irrelevantes? Nossas convicções materialistas do século XXI tornam difícil compreender o que seriam objetos que não produzem efeito no mundo físico. Em outras palavras, postular a existência de algo parece implicar o reconhecimento de suas relações causais com outros objetos ou eventos físicos.

⁶⁸ Aqui estarei adotando uma concepção ampla de fisicismo que engloba tudo aquilo que pode ser objeto das ciências da natureza.

A segunda alternativa é reducionista ou eliminativista, ou seja: ou bem reduzimos os objetos fenomenais aos objetos físicos, negando suas características peculiares, ou negamos a existência de tais objetos e como consequência o reconhecimento de propriedades *sui generis*. Esta alternativa deixa intacto o fisicismo, porém não nossas práticas cotidianas de atribuição de poderes causais a nossos estados mentais. A terceira alternativa parece compatível com o fisicismo, com um dualismo ou pluralismo de propriedades e com certa compreensão do caráter peculiar de algumas de nossas vivências.

Tanto a segunda quanto a terceira alternativa compartilham o núcleo das intuições fisicistas e a diferença entre ambas é bem mais sutil do que pode nos parecer à primeira vista. Algum tipo de reducionismo ou mesmo de eliminativismo parece inevitável até mesmo à terceira alternativa. A diferença entre ambas consiste apenas na tentativa desta última de conciliar o fisicismo com algumas de nossas mais caras intuições acerca de nossos estados subjetivos. Mas que estados subjetivos especificamente? Ou melhor, que características peculiares aos mesmos parecem razoáveis se sustentarem junto ao fisicismo, ou seja, sem que tenhamos que abandonar o núcleo de nossas convicções fisicistas?

Parte da solução consiste na caracterização do que seria o *explanandum* do problema. O que parece oferecer resistência a uma redução fisicista não são os estados mentais em geral, mas certas propriedades que supomos pertencer a estados mentais específicos. São propriedades como: privacidade, necessário pertencimento a um sujeito e perspectiva. Tais propriedades parecem exprimir o caráter subjetivo da experiência, ou seja, se reportam ao que é vivenciado única e exclusivamente na perspectiva do sujeito da experiência. Trata-se do conjunto de vivências para as quais Nagel introduziu a expressão “what it is like to be” (Nagel 1991). Na perspectiva de Nagel, tais propriedades acompanhariam todo o tipo de evento intencional, ou seja, todo tipo de relação que envolve a consciência e seus objetos, constituindo o caráter subjetivo, irredutível das mesmas.

Meu primeiro passo rumo a uma descrição funcional do mental será o de defender uma posição bem mais estreita frente à chamada consciência fenomenal: uma posição que, ao contrário de Nagel, exclui deste âmbito todos os estados mentais cujo conteúdo relevante pode ser descrito como uma atitude proposicional. Este é o caso de crenças, desejos e intenções etc. Para justificar tal restrição pretendo mostrar que é possível fornecer, ao menos *prima facie*, uma descrição funcional de nossas atitudes proposicionais e que tal descrição parece não eliminar ou excluir nenhum elemento relevante para a compreensão das mesmas.

O irredutível caráter subjetivo da experiência - que compreendo pelo título de *qualia* - será encarado como uma qualidade peculiar dos estados mentais relativamente aos quais a perspectiva de quem os vivencia parece acrescentar algo de irredutível à própria experiência. Com isso pretendo caracterizar o problema da lacuna explicativa como dirigido especificamente aos estados mentais que parecem

não poder ser descritos, satisfatoriamente, sem uma referência direta àquele que os vivencia ou, em outras palavras, a suas qualidades fenomenais. Em seguida, pretendo analisar o lugar dos *qualia*, agora entendidos em sentido estrito, em uma perspectiva funcionalista da mente.

A perspectiva funcionalista, em geral, caracteriza-se por identificar os estados mentais à função realizada por estados cerebrais ou ao estado cerebral que realiza esta função. O que torna um estado cerebral uma crença de conteúdo *p* é a sua relação com determinados estímulos sensoriais, com outros estados internos e com o comportamento. Neste sentido, os estados mentais são tanto causas internas do comportamento, como efeitos produzidos pelo mundo externo. Nas palavras de Jackson e Braddon-Mitchell:

(...) a teoria funcionalista da mente especifica estados mentais em termos de três tipos de cláusulas: as cláusulas de entrada que especificam que condições dão origem, tipicamente, a que estados mentais; cláusulas de saída que especificam que estados mentais dão origem, tipicamente, a que repostas comportamentais; e cláusulas de interação que especificam como estados mentais interagem tipicamente. (Braddon-Mitchell e Jackson 1996, p.40)

Para que não haja circularidade na explicação, a caracterização de cada etapa deve ser realizada em termos meramente físicos, ou seja, os estados intencionais que pretendemos descrever em termos físicos não podem ser reintroduzidos na caracterização do *output*. Os *outputs* devem poder ser totalmente descritos em termos de alterações físicas no mundo. A descrição funcional de estados mentais envolve, assim, uma rede de relações físicas que inclui disposições comportamentais, causas típicas e outros estados mentais, desde que também caracterizados em termos de disposições comportamentais, ou seja, em termos funcionais.

A forma mais usual de exemplificar o funcionamento da teoria é através da construção de modelos, máquinas programadas para realizar um tipo de funcionamento específico. Neste caso, as ilustrações vão desde máquinas mais simples como as que nos oferecem Coca-Cola mediante a introdução de moedas de um determinado valor, até as máquinas conexionistas de tipo PDP (Parallel Distribution Process) ou redes neurais. Parece evidente que a complexidade das funções realizadas por nossos estados mentais exige modelos flexíveis e bastante complexos. Complexos talvez o suficiente para não sermos, ainda, capazes de descrevê-los. Dessa forma, a proposta funcionalista parece vulnerável a uma objeção de fato, mas, se estivermos corretos, poderemos mostrar que, em princípio, nada impede que todos os aspectos relevantes do mental possam ser descritos em termos físicos.

A mente humana, descrita funcionalmente, poderia, assim, operar como um programa flexível composto de vários módulos. No primeiro módulo estaria um *scanner*, responsável pela recepção dos *inputs*. A partir daí, podemos imaginar vários

módulos entre os quais um módulo avaliador, responsável pela seleção das informações que chegaram à etapa final, qual seja, a produção de um comportamento específico. A peculiaridade deste tipo de programa estaria na sua capacidade de aprender, ou seja, de alterar o produto em função de um novo *input* que incluiria os efeitos produzidos pelo comportamento do agente nas etapas anteriores. Em outras palavras, o *output* gerado promoveria respostas externas que, por sua vez, seriam introduzidas no *input* e avaliadas pelo programa de forma a fornecer um novo resultado. Deste modo, um modelo programado para reconhecer letras, como o que é utilizado com sucesso nos correios, altera, ou melhor, aprimora, sua *performance*, na medida em que lhe são oferecidas grafias variadas de uma mesma letra e que o programador responde negativa ou positivamente ao seu reconhecimento ou não de uma letra. Tais modelos são capazes de reordenar os seus dados de forma a passar a reconhecer características anteriormente ignoradas. Com base na análise funcional de modelos com programa flexível passaríamos, então, a atender uma das principais características do modelo humano: sua capacidade de aprendizado.

Supondo, assim, que uma descrição funcional de nossas atitudes proposicionais seja possível e que com tal explicação nada de essencial às mesmas se perca, o problema da lacuna explicativa ou o problema da redução fisicista passa a se concentrar exclusivamente na questão dos *qualia*. Como encaixá-los em uma descrição física do mundo? Ou ainda, a que exatamente tais propriedades se referem?

Se não estivermos dispostos a aceitar uma perspectiva reducionista ou eliminativista a respeito dos *qualia*, devemos então ser capazes e explicar em termos funcionais como as propriedades “*what-it-is-like*” podem desempenhar um papel causal genuíno em nossas disposições comportamentais. Para tal, pretendo me servir agora da teoria PANIC desenvolvida por Michael Tye. Nas palavras do autor: “phenomenal character is one and the same as Poised Abstract Nonconceptual Intentional Content.”⁶⁹

A teoria PANIC é uma teoria sobre os eventos ou estados fenomenais que os reconhece como fenômenos intencionais, porém, ao mesmo tempo, como fenômenos não-conceituais. Tais estados podem acompanhar outros estados intencionais de conteúdo proposicional e podem servir de *background* para processos cognitivos envolvendo estados intencionais de conteúdo proposicional. O decisivo, porém, é que os estados fenomenais intencionais *não* são eles mesmos conceituais. Sensações corporais como, por exemplo, a dor de dente ou o orgasmo são representacionais no sentido em que refletem causalmente, *em circunstâncias normais*, as alterações em partes específicas do corpo. Estas alterações formam o conteúdo representacional das sensações corporais. Tais representações (i) são o *output* de receptores nervosos respondendo a estímulos diversos no nosso corpo e (ii) integram o *input* dos

⁶⁹ Tye, M., *Ten Problems of Consciousness*, Cambridge, Mass.: Bradford Books, 1995, p.137.

processos cognitivos em que são, pela primeira vez, reveladas cognitivamente ao sujeito das representações. Mediante a ação de processos cognitivos operando sobre as representações não-conceituais, o sujeito se torna, então, consciente de (*aware of*) seus estados fenomenais.

Tornar-se consciente de (*to became aware of*) suas próprias experiências é um processo cognitivo que subsume, pela primeira vez, a experiência fenomenal a conceitos. Só então o sujeito experiencial adota uma atitude epistêmica (cognitiva) frente a seus estados fenomenais. *Abaixo* deste nível de *awareness* -chamado tradicionalmente de “introspecção” - está a consciência fenomenal (*consciousness*). O conteúdo fenomenal seria, portanto, mais bem descrito como representações não-conceituais de alterações físicas prontas a integrar o *input* intencional de processos que têm como *output* outros estados intencionais de conteúdo proposicional (*beliefs* e *desires*). Ser o sujeito de uma experiência fenomenal significaria assim, ter, no *input* de certos processos cognitivos, representações adicionais, representações não-conceituais que poderiam promover respostas diferenciadas.

Aceita a descrição do mental nos termos aqui propostos, mente e corpo já não poderão mais ser pensados como entidades distintas. A pergunta acerca do que somos deverá agora ser respondida por referência a uma rede de processos que envolvem a *performance* de distintas funções, algumas das quais usualmente descritas através de um vocabulário mentalista. A esta rede passarei a chamar Self. O Self, assim entendido, não é uma unidade transcendente que controla todo o sistema, nem uma parte específica do mesmo. Ele é uma rede ou uma conjunção de processos. Enquanto tal, ele está projetado no mundo e em constante processo de transformação. Seu campo informacional é composto de dados oriundos tanto dos limites internos, quanto externos ao próprio corpo. Nossa memória, por exemplo, não contém apenas informações registradas por um processo de introspecção ou por um exame de nosso cérebro ou sistema neurológico. Ela envolve também os diversos objetos e processos a partir dos quais expandimos nossas mentes. Os registros deixados em livros, diários, computadores e no nosso meio-ambiente ocupam um papel fundamental em nosso *set* informacional.⁷⁰ Este fato talvez explique porque nos sentimos perdidos quando perdemos nossa agenda ou as informações de colocamos em nossos computadores. Pode explicar também a experiência intensa que vivenciamos ao retornar a lugares percorridos na infância ou a perda de capacidades mentais em idosos que são afastados do lugar e das coisas com as quais conviveram durante muitos anos. O Self não é uma entidade que se relaciona ou representa o mundo. Ele é uma rede de processos no mundo.⁷¹

⁷⁰ A ideia de Self externalizado é desenvolvida em Daniel Dennett, *Kinds of Minds: Toward an Understanding of Consciousness*, Science Master Series, 1996.

⁷¹ Sobre este ponto ver: W. Teed Rockwell, *Neither Brain nor Ghost*, Cambridge, Mass: The MIT Press, 2005.

2. Do Self funcional ao agente moral

Dito isto passamos então à questão fundamental: Pode uma descrição funcional dos eventos mentais e, conseqüentemente, do Self satisfazer a nossa concepção comum do que seja um ser humano, uma pessoa,⁷² e preservar o caráter peculiar que lhe é atribuído como agente e/ou sujeito moral?

Como decorrência da adoção de uma perspectiva naturalista, seres humanos são então aqui compreendidos como um sistema sócio-biológico que se constitui como parte do meio-ambiente e que se transforma de acordo com o mesmo. Mas será que assim seremos capazes de distingui-los de outros sistemas e resgatar os privilégios até hoje requeridos? Proponho que antes de tentarmos responder a esta pergunta consideremos algumas das fórmulas tradicionalmente fornecidas na filosofia para caracterizar e distinguir o status concedido aos seres humanos.

Somos, antes de tudo, compreendidos como seres racionais.⁷³ Mas o que isto realmente significa? Para tentar responder a esta questão de forma compatível com a boa tradição filosófica e, ao mesmo tempo, compatível com nossas convicções pós-metafísicas, pretendo analisar a explicação do que seja racionalidade fornecida por Joseph Raz.⁷⁴ No sentido mais trivial Raz caracteriza os seres racionais como todos os que possuem “capacidade de racionalidade” (‘capacity-rationality’). Esta capacidade é descrita como a habilidade para perceber razões e agir conforme as mesmas. A capacidade de racionalidade envolve uma série de outras capacidades, tais como a habilidade perceptiva, a posse de memória, a capacidade de pensamento conceitual, a capacidade de formar opinião e tomar decisões e a capacidade de controlar os próprios movimentos. Alguém que careça de um grau mínimo de alguma destas habilidades não satisfaz as condições para o exercício da racionalidade e, nesse sentido, suas ações não podem ser ditas racionais ou irracionais, pois ao agente faltaria a própria capacidade de racionalidade. Tal como as demais habilidades mencionadas, a capacidade de racionalidade será compreendida como algo gradativo. Alguém poderá ser reconhecido como mais ou menos racional de acordo com sua *performance*, tanto nas habilidades pressupostas pela racionalidade, como no exercício da própria racionalidade. A capacidade da racionalidade não deverá ser identificada nem com o processo de deliberar ou decidir, nem como resultado de

⁷² Estou utilizando aqui o termo “pessoa” para evitar qualquer tipo de especismo, i.é. a ideia de que somos especiais simplesmente porque pertencemos a certa categoria: à espécie humana.

⁷³ Ser racional para muitos filósofos, tais como Kant, Habermas ou Rawls, envolve um comprometimento com justificações morais e ou com o agir de acordo com princípios morais. Para estabelecer tal conexão, é necessário uma complexa teoria da racionalidade que não pretendo discutir aqui.

⁷⁴ Ver, Joseph Raz, *Engaging Reason: on the theory of value and action*. Oxford: Oxford University Press, 1999.

algum desses atos. Ela se exprime automaticamente no modo como funcionamos, quando razões são capazes de afetar nossas crenças, desejos e ações.⁷⁵

Essa seria então, na interpretação de Raz, a capacidade que nos distinguiria e nos caracterizaria enquanto pessoas. Em suas palavras:

*Our capacities to perceive and understand how things are, and what response is appropriate to them, and our ability to respond appropriately, make us into persons-creatures with the ability to direct their own life in accordance with their appreciation of themselves and their environment (...)*⁷⁶

Como essa concepção Raz recusa a distinção tradicional entre razão teórica e razão prática⁷⁷ e a distinção mais recente entre razão procedural e razão substantiva.⁷⁸

Até aqui, nada em sua concepção de racionalidade é incompatível com uma descrição funcional do Self. Racionalidade é apresentada como uma capacidade ou uma função desenvolvida em continuidade com outras capacidades ou funções. Restaria esclarecer o que podemos considerar razões e como podemos ser afetados por elas. Não pretendo, contudo, discutir a concepção de razão em Raz. Para meu propósito é suficiente que compreendamos razões como dados informacionais que possam aprimorar a habilidade do Self de responder adequadamente a diferentes inputs. O Self aprende a identificar as informações mais importantes e a reagir com base nas mesmas. Isto é o que queremos dizer com “ser sensível a razões”. Para resumir, poderíamos chamar “racional” os seres ou sistemas que possam se comportar com base na seleção e avaliação de informações. Quão complexo e desenvolvido é este processo de avaliação irá determinar o grau de racionalidade do sistema.

Mas se falamos aqui livremente de seres ou sistemas será que realmente chegamos a uma descrição satisfatória para o que supomos ser uma pessoa? Algumas vezes chamamos de pessoa, seres que não podem, ou que nunca poderão, responder apropriadamente ao seu entorno (meio); seres incapazes de desenvolver de forma satisfatória as capacidades anteriormente mencionadas. Outras vezes, atribuímos certo grau de racionalidade não apenas a animais, como também a computadores: às assim chamadas “máquinas inteligentes”. Se o primeiro caso for realmente verdade, então deveríamos dizer (1) que a capacidade da racionalidade não é necessária para identificar uma pessoa ou (2) que nosso uso comum da palavra

⁷⁵ Ver, idem, ibidem, 69.

⁷⁶ Idem, ibidem, 67.

⁷⁷ Tal como proposta por Kant.

⁷⁸ Ver, D. Parfit, “Reason and Motivation”. *Aristotelian Society*, suppl. 71 (1997), 99. Contra a posição de Parfit, ver: Raz, 72-73.

“pessoa” ultrapassa o uso legítimo deste conceito. Se o segundo caso de fato ocorre, então precisaremos (1) investigar outra característica distintiva para os seres humanos/pessoas ou (2) aceitar que somos apenas mais um sistema racional no mundo. Antes de decidirmos por alguma destas posições, proponho que investiguemos outras possíveis descrições do que seja uma pessoa.

Podemos chamar de racional um animal ou uma máquina, mas não nos sentiríamos suficientemente à vontade para atribuir-lhes liberdade ou responsabilidade por suas ações. Por que não? Porque pensamos que isso envolveria algo mais. Algo reconhecido apenas nas ações humanas: algo que supomos expresso no conceito de liberdade da vontade.

Em seu famoso artigo “Freedom of the Will and the Concept of a Person”⁷⁹ Harry Frankfurt defende que a distinção essencial entre pessoas e outros sistemas está na estrutura da vontade. Pessoas não são para ele apenas aqueles aos quais atribuímos estados mentais, que possuem desejos e são capazes de tomar decisões, mas sim aqueles que possuem a capacidade de constituir desejos de segundo nível.⁸⁰

Seguindo sua concepção, falamos de (i) desejos de primeiro nível quando alguém busca realizar aquilo que quer e de (ii) desejos de segundo nível quando quer ter certos desejos ou quando quer que certos desejos se imponham como a sua vontade. No segundo caso, falamos, então, de volições de segundo nível. A característica essencial de uma pessoa é fornecida para Frankfurt não apenas pela presença de desejos de segundo nível em geral, mas pela presença de volições de segundo nível. Porque uma pessoa tem volições de segundo nível, ela é igualmente capaz de desfrutar de uma vontade livre.⁸¹

Mas o que exatamente significa falar de níveis de desejos e de liberdade da vontade? De acordo com nosso uso mais comum do conceito de liberdade, alguém é livre quando realiza aquilo que quer ou decide realizar.⁸² Contra esta definição trivial de liberdade, Frankfurt apresenta o conceito de liberdade da vontade. Ser livre neste sentido, não é ser livre para fazer o que se quer, mas sim para desejar o que se quer.⁸³ Mas o que significa dizer que alguém é livre para desejar o que quer? O que está realmente implicado pelo conceito de liberdade da vontade que falta ao conceito comum de liberdade?

⁷⁹ Frankfurt, H., “Freedom of the Will and the Concept of a Person”, *The Journal of Philosophy* 68 (1971).

⁸⁰ Ver: idem, ibidem, 6.

⁸¹ Ver: idem, ibidem, 11-14.

⁸² Ver: idem, ibidem, 14.

⁸³ Ver: idem, ibidem, 18-19.

Deixando de lado especulações metafísicas acerca da liberdade, podemos dizer que a atribuição ou não de liberdade a alguém tem consequências práticas inquestionáveis. De acordo com o senso comum, só responsabilizamos alguém pelas consequências de seus atos, se o supusermos livre para agir de acordo com sua própria decisão. Mas será que para garantir a plausibilidade de tais práticas precisamos ir além do que do nosso conceito comum de liberdade? Eu penso que não. Para reconhecer alguém como responsável pelos seus atos, não precisamos imaginar diferentes níveis de desejos, nem nos comprometer com algum tipo de premissa acerca da liberdade da própria vontade. Liberdade como capacidade de determinar as próprias ações pode ser entendida, em sentido trivial, como a capacidade de agir baseado na avaliação de crenças e desejos.⁸⁴

Seguindo esta posição, dizemos que alguém é responsável por seus atos quando reconhecemos que ele poderia ter agido de outro modo, caso assim decidisse. Em outras palavras, quando reconhecemos que agiu sem coação e que foi capaz de escolher entre alternativas distintas. Ausência de coação e a presença de alternativas fazem, portanto, parte do *background* a partir do qual julgamos alguém como livre e, por conseguinte, como responsável por suas ações.

No entanto, podemos imaginar situações nas quais o agente é livre de coação, age com base em seus desejos, mas ainda assim não estamos inclinados a dizer que ele foi responsável pelos seus atos. Este é o caso de situações em que o agente em questão são animais, crianças pequenas ou mesmo adultos carentes das habilidades necessárias ao exercício da racionalidade. Assim, podemos indicar a anteriormente mencionada capacidade da racionalidade, i.é, a capacidade de agir baseada na avaliação das crenças e desejos mais relevantes, como uma condição necessária, embora não suficiente, para a atribuição de liberdade. A capacidade de determinar as próprias ações de acordo com suas próprias metas é o que chamamos liberdade como autodeterminação. Um ser autodeterminado é aquele capaz de agir com base em razões: com base em um *set* relevante de informações acerca de suas crenças, seus desejos e seu entorno (meio-ambiente, *background*).

⁸⁴ Tal capacidade é o que chamamos algumas vezes de capacidade de deliberar. Neste artigo tentei intencionalmente evitar o uso do conceito de deliberação. A razão disto é porque, tal como eu entendo, deliberação é um processo: a totalidade do processo de verificar, selecionar e avaliar informações e reagir. Deste modo, ao falarmos das diversas etapas do processo, já estaremos falando sobre a deliberação e não seria necessário introduzi-la como um elemento a mais ou a parte. Sei que muitos autores costumam identificar deliberação com apenas uma parte do processo: a etapa entre a avaliação ela mesma e a resposta. Eles a pensam como um momento de decisão que poderia ser discriminado separadamente. Tal ideia me parece absurda. Eu simplesmente não posso compreender o que possa ser uma decisão em si mesma, sem pensar em seu conteúdo - i.e. as informações e alternativas em questão - e no meio pelo qual é expressa, i.e. a ação. Qualquer outra alternativa me parece uma reminiscência da concepção cartesiana.

Em muitos contextos, o que caracteriza uma pessoa e permite distingui-la de outros sistemas pode ser resgatado através do conceito de liberdade como autodeterminação. Mas o que este conceito de liberdade realmente acrescenta ao conceito de ser racional anteriormente mencionado? Do ponto de vista da descrição funcional, eu diria: nada. A diferença aparece apenas no nível do *background*. Só faz sentido falar de autodeterminação quando não há coação externa, em outras palavras, quando o resultado do processo pode ser determinado internamente. Para garantir isto, autores como Kant e Frankfurt introduziram a noção de uma vontade soberana, alheia a todas as motivações sensíveis ou desvinculadas das leis causais do mundo natural. Contudo, se seguirmos a descrição funcional aqui apresentada não é difícil compreender como os diversos estágios de um processo podem gerar autonomamente uma resposta. Usando o exemplo trivial da máquina que nos fornece Coca-Cola, poderíamos dizer que o processo autonomamente gerou uma resposta, quando não chutamos a máquina para conseguir a lata de Coca-Cola ou desligamos sua fonte de energia. Em casos mais complexos, nós atribuímos o poder de autodeterminação quando o *background* nos permite a exclusão de fatores que possam gerar um desequilíbrio radical nos pesos atribuídos aos dados do sistema. Fatores deste tipo são o que chamaríamos coação, quer atuem fora ou dentro dos limites corporais do sistema. No exemplo de ações humanas típicas, nosso *background* exclui a atribuição de liberdade tanto em casos de graves distúrbios internos, quanto no de extrema violência externa.

3. Consequências morais e políticas da adoção de uma perspectiva naturalista: O ideal do pleno desenvolvimento dos funcionamentos

Teremos, então, alcançado uma caracterização do ser humano suficiente para distingui-lo de máquinas e outros animais? Minha resposta é: sim e não. Sim se estivermos dispostos a aceitar que o que nos distingue sejam apenas o grau e a complexidade com que realizamos algumas funções. Não, se insistirmos na crença de que algo de substancial pode marcar a diferença da nossa forma de existência no mundo. Não creio que para dar conta das questões que envolvam nossas atitudes perante seres humanos, seja necessário prosseguir nessa busca. Mas se aceitamos que toda a diferença está apenas do grau de desempenho de certas habilidades, então podemos imaginar um novo quadro no futuro e devemos estar abertos para a possibilidade de expandirmos a mesma consideração ou respeito que exigimos hoje frente aos atuais "seres humanos" a outros sistemas. Este, no entanto, não parece ser um problema, pois não apenas já ampliamos, relativamente ao passado, o âmbito dos que merecem a nossa consideração (ou aqueles aos quais atribuímos certos direitos), como, também, a cada dia, ampliamos o nosso discurso moral de forma a abarcar também o respeito pelas demais formas de vida. Um exemplo disto é a defesa dos direitos de terceira geração.

Mas o que fazer com os ditos seres humanos que não se enquadram na descrição fornecida? Ou seja, que não possuem a capacidade de autodeterminação e que apenas apresentam um grau bem débil das demais habilidades mencionadas? Esta parece ser a questão mais relevante sob o ponto de vista das discussões atuais em bioética. Longe de tentar aqui respondê-la de forma exaustiva, pretendo apenas chamar atenção para a necessidade de algumas revisões. Uma descrição funcionalista do Self tem como consequência uma desmistificação da ideia de ser humano/pessoa.⁸⁵ Neste sentido admitir que certos sistemas não caracterizam mais um ser humano não é um sacrilégio. Sob o ponto de vista moral, a consequência mais imediata é uma desantropologização da ética. Não interessa se certas criaturas são ou não humanas. O que interessa é que possamos agir de tal forma a sempre promover o melhor funcionamento possível dos diversos sistemas. Devemos aprimorar nossa compreensão dos diversos funcionamentos e suas respectivas demandas, visando uma ampliação dos nossos padrões de qualidade de vida.

No âmbito da saúde⁸⁶ e da política⁸⁷ index de Qualidade de Vida já vem sendo propostos e utilizados como guia de conduta. O que é um tratamento adequado em cada caso irá depender do conjunto de capacidades envolvidas, i.é, das funções que podem ser desempenhadas por cada sistema. Dentro deste universo funcional incluímos funcionamentos bem simples – tais como a capacidade de alimentar-se e a vulnerabilidade ao sofrimento – como também capacidades bastante complexas, tais como a de engajar-se em uma sociedade política, de educar filhos, pintar ou escrever livros. Baseados unicamente em uma descrição funcional dos objetos de nossa consideração moral não teremos razões para discriminar certas formas de vida e supor serem algumas mais valorosas ou dignas de respeito. Que tipo de conduta melhor expressa nosso respeito pelos demais irá variar de um contexto ao outro. Deixar um sistema morrer pode ser interpretado tanto como um caso de indiferença moral, por conseguinte, algo condenável moralmente, como a expressão do quanto estamos atentos ao sofrimento alheio e reconhecemos, até o último instante, o direito eminentemente humanos à autodeterminação. É, portanto, hora de rever nossas

⁸⁵ Pretendo aqui recusar os preconceitos embutidos tanto na noção de ser humano quanto de pessoa, i.é. tanto o “humanismo” quanto o “personalismo”. Acerca desta disputa entre humanistas e personalistas, ver: P. Singer, *Practical Ethics*, Cambridge: Cambridge University Press, 1993 e B. Williams, “The Human Prejudice”, in B. Williams e A. W. Moore (eds.) *Philosophy as a Humanistic Discipline*, Princeton: Princeton University Press, 2006.

⁸⁶ Um exemplo é o The Sickness Impact Profile (SIP) desenvolvido por Marilyn Bergner e sua equipe, o the Quality of Life Index (IQL) desenvolvido por Walter O. Spizer e equipe e o Health Status Index (HIS) desenvolvido por Milton Chen e equipe. Sobre este tema: Dan Brock, “Quality of Life in Health Care and Medical Ethics”, in: M. C. Nussbaum and A. Sen (eds.) *The Quality of Life*, Oxford: Oxford University Press, 1993.

⁸⁷ Tal ideia é desenvolvida por Amartya Sen e Martha Nussbaum em vários artigos. Uma síntese desta discussão é encontrada no livro *The Quality of Life*.

convicções e tornar nossa perspectiva moral mais inclusiva e mais compatível com tudo o que temos aprendido acerca de nós mesmos e de nosso entorno.

Mind and Person in a Physical World⁸⁸

From ancient to contemporary days, philosophy has been trying exhaustively to explain human's performance in the world. Our role, not only as producer and subject of knowledge, but also as moral agent and reality transformer, has been always, direct or indirectly, related to some aspects of the human mind. The experience of a privileged access to our own mental state made our identification while *res cogitans* as the only certainty in a world in which a malignant genius used all his power to deceive us. We are thus identified, first of all, with the set of our mental state: thinking, doubting, desiring etc. Just after a long and tenebrous empire of the hyperbolic doubt, we have also rescued our identity as *res extensa*. From a mind only, we started to be identified as mind and body. So a problem that occupies philosophers in several areas until now was established: what is the relation between these two parts, mind and body, with which we should be identified? A satisfactory answer to this question should not just clarify what we are but also justify the status that we have been complaining in the world. In this article, I intend to defend a physicalist perspective of mind and, therefore, a materialistic conception of self. Within this perspective, I will defend that we cannot identify what we used to attribute mental properties as either a nonphysical entity, or with the brain. So I intend to reject the mind-body dualism, so much in its cartesian version, as in the contemporary materialistic version. Mental properties, as well as other physical properties, are attributed to systems and the functions accomplished by them. To defend such perspective, I intend, first of all, to explain under the functionalist perspective, our intuitions concerning the peculiar character of mental events. After that, I intend to show how a physicalist conception of self can fit our common sense comprehension of person.

We can understand physicalism, in general, as the perspective according to which everything that exists or that may be known belongs to the physical world, i.e. to the domain of objects and laws of an elementary microphysics of the future. Whichever definition of the “physical world” we would adopt, it would seem that it should incorporate a set of properties applicable in fact to a specific group of objects that, at least on a first impression, cannot be imputed meaningfully to physical objects. For instance, we say that “no one else can feel my pain” and also that pain, sadness, gladness and other mental states just exist while lived by a subject or while belonging to a subject. My consciously perceptible experience, my body sensations,

⁸⁸ This article was written during my stay as academic visitor with the Uehiro Centre for Practical Ethics at Oxford University. I would like to thank especially Julian Savulescu, with whom I worked at the University of Oxford, Neil Levy and Nicholas Shean for their careful comments. I also had the opportunity, for which I am grateful, to discuss the main issues presented here with Carlos Eduardo Evangelistis Mauro and Sofia Miguens, during a conference in Porto.

emotions, passions and moods seem to be related to me in a way that my car and my clothes, for instance, cannot be. Another person could have my car and my clothes and those items would not cease to exist if not possessed by someone.

The “monopolization” of experienced mental phenomena by a subject seems to characterize a sort of necessarily inner belonging, which is alien to the objects that only contingently belong to me. Should we not distinguish the objects to which I have a privileged access, that is, the objects that can only be “possessed” and “known” in my experienced perspective, and that cannot exist unless they are related to the subject of the experience from the other objects to which I simply relate to, because I am the one that contingently possesses them? If the characteristic properties of the objects of first group - that is, the property of privacy, the property of being necessarily owned and the property of perspectivity⁸⁹ - cannot be attributed to physical objects, should we not then admit the existence of nonphysical entities?

That dualist conception seems to violate a widely accepted principle that the physical world is causally closed. This principle says that for any physical event whatsoever its causes are other physical events. Nonphysical entities have, therefore, no causal role to play in the processes by which physical events are brought about. Phenomenal events could of course be still effects, but never causes of physical events. This is the epiphenomalist solution to problem of the mental events. But are we really prepared to admit that mental states are forever causally irrelevant?

Many people share the intuition that there is a gap between the physicalist explanation of our mental life and what we innerly believe to happen in some contexts. Would it be possible to explain, in physicalist terms, that is, to say, in terms of elementary microphysics, even if only in principle, what it means to be the subject of some experiences, to be the agent of our actions, or to express the privileged point of view that we assume towards our mental states, sensations, emotions, passions, and moods?

Faced with this problem, we may find three alternatives:

1. to accept the *sui generis* character of phenomenal objects and their irreducibility to the domain of the physical world;
2. to show that phenomenal objects do not really have these *sui generis* properties; or to show that these objects do not exist and, thus, that we cannot apply these properties; or

⁸⁹ On these properties, see: Michael Tye, *Ten Problems of Consciousness*, Cambridge, Mass.: Bradford Books, 1995.

3. to acknowledge the existence of phenomenal objects and their characteristic properties denying, however, that they do not apply to physical objects;

The first alternative is clearly dualist. It reinstates an ontological dualism in which the relation between the mental and the physical becomes contingent, arbitrary or even impossible. At best, it would be possible to restore an epiphenomenalism, according to which the mental would be causally inert. Our 21st century materialist beliefs, however, make it difficult to understand what it means to think about objects that do not produce effects on the physical world. How could these objects ever be perceived if they do not leave traces in the physical world? To claim the existence of something seems to imply causal relations with other physical objects or events.

The second alternative is either reductionist or eliminativist. Respectively, it either reduces the phenomenal objects to physical ones, thereby denying their peculiar characteristics; or it denies the existence of these objects and consequently the realization of *sui generis* properties. This alternative leaves physicalism untouched, but not our ordinary practices of attributing causal powers to mental states.

The third alternative seems to combine physicalism with some kind of dualism (or pluralism) of properties, thus affording some understanding of the peculiar character of some of our experiences. The second and the third alternatives share the kernel of the physicalist intuition. The difference between them is subtler than it seems to be at first sight. Some sort of reductionism or eliminativism seems unavoidable even to the third alternative. The difference between them lies only in the attempt of the third alternative to conciliate physicalism with some of our most valuable intuitions about our subjective states.

To solve the problem of the peculiar character of our mental states and to conciliate our conception of the mind with our materialist convictions about everything that exist, I intend to develop the third alternative and give a functionalist account of mental events.

The functional description of mental states specifies a net of physical relations involving dispositions to behavior, typical causes, as well as other mental states as far as these are also characterized in functional terms. What grounds the attribution of a certain content p to a state of belief, for instance, is the relation of this state with certain sensorial stimuli, with other inner states, and with behavior. Therefore, mental states are both inner causes of behavior and effects produced by the external world. In the words of Jackson and Braddon-Mitchell:

A functionalist theory of mind specifies mental states in terms of three kinds of clauses: input clauses which say which conditions typically give rise to which mental states; output clauses, which say which mental states typical give rise to which

*behavioral responses; and interaction clauses, which say how mental states typically interact.*⁹⁰

To avoid circularity, the functionalist characterization of any intentional state must be made exclusively in nonintentional, physical terms. In particular, the propositional attitudes that we intend to describe in physical terms must not be surreptitiously reintroduced in the characterization of the output. As with inputs, outputs must also be described entirely in physical terms.

The most obvious applications of the theory follow the model of a machine programmed to perform specific tasks. This includes very simple machines, like vending machines that provide us with Coke when we introduce a coin of a certain value, but it also includes highly complex ones like connectionist machines of the PDP kind or neural nets. Evidently, the complexity of the functions realized by mental states requires flexible and very complex models. In many cases the complexity is so as that we are as yet unable to specify the models in question. This, however, is no objection to the claim that, in principle, there is nothing to prevent relevant aspects of mental events from being described in physical terms.

The human mind described in a functional way thus operates like a connectionist machine composed of several modules. There would be a scanner, which is responsible for the reception of the inputs. From that point on, we can imagine several modules. Among them, there would be an evaluator module, responsible for the selection of the pieces of information generated by an alternative generator. Finally, some motor module outputs a specific behavior. The peculiarity of such an assembly of modules is its capacity to learn, that is, to alter the final result, due to new inputs reflecting the effects produced by the agent's behavior in the previous stages. In other words, the produced output would promote external answers, which, in their turn, would be introduced as input and evaluated by the program, so that it can provide a new result. Thus, e.g., a program able to recognize alphabetic characters, like the ones successfully used in postal services, transforms and improves its performance when different tokens of the same character are provided and the manager answers affirmatively or negatively to its recognition. Such programs are capable of organizing the information so that a character previously misinterpreted can later be recognized.

So, based on the functional analysis of flexible program models, we can explain the most typically characteristics of the human cognitive and emotional system: the capacity of selecting information, evaluating, reacting and learning.⁹¹

⁹⁰ Braddon-Mitchell, D. and Frank Jackson. *The Philosophy of Mind and Cognition*. Oxford: Blackwell, 1996, p. 40.

⁹¹ For a defense of that model, see Blackburn, S. *Think*. Oxford: Oxford University Press, 1999, p. 98f.

Can we really say that we have already solved the problem? Actually, what seems to resist to a physicalist reduction are not mental states in general, but some of their properties: privacy, necessary ownership and perspectival subjectivity. These properties reflect the subjective character of our experience; they are connected to what is lived only and exclusively from the perspective of the experienced subject. To use Nagel's phrase, those properties constitute the "what it is like to be"⁹² character of our experience. In Nagel's view, these properties are present in every intentional event, i.e., in every kind of relation that would involve consciousness and its objects. They constitute the irreducible subjective character of intentional events.

Against Nagel, I will sustain a more plausible and restricted position about the phenomenal consciousness, a position that excludes from this domain our mental states such as beliefs, desires and intentions whose relevant content is expressed by a that-clause. The irreducible subjective character of experience, or *quale*, will be seen as a peculiar quality not of all but of some specific mental events: the ones that do not seem to be sufficiently described without a direct reference to the agent. Due to these peculiar qualities, the perspective of the experiencing subject seems to add something irreducible to the experience itself; something that cannot be described in conceptual terms. That peculiar subjective character of our phenomenal consciousness is in fact what characterizes the problem of the explanatory gap. To solve the explanatory gap, we need to explain how the phenomenal character of consciousness, *qualia*, also fits in a physical description of the world. In other words, we must explain, in functional terms, how the what-it-is-like properties may play a genuine causal role in our behavioral dispositions.

In developing this possibility I draw on Michael Tye's PANIC theory of phenomenal character: "phenomenal character is one and the same as Poised Abstract Nonconceptual Intentional Content."⁹³ Though Tye's theory distinguishes between the nonconceptual aspect from the abstract one, I will consider exclusively the nonconceptual side of phenomenal experience, which is the relevant one for the present purposes.

The PANIC theory is a theory of phenomenal events or states as intentional but nonconceptual phenomena. These states may come together with other intentional states with propositional content. They can work as background elements to cognitive processes involving other mental states. The crucial point here is that the intentional phenomenal states, as opposed to propositional attitudes, are not themselves conceptual. Bodily sensations, such as a toothache or an orgasm are representational,

⁹² Nagel, T. "What it is like to be a bat," in: *Mortal Questions*. Cambridge: Cambridge University Press, 1979.

⁹³ Tye, M., *Ten Problems of Consciousness*, Cambridge, Mass.: Bradford Books, 1995, p.137.

in such a way that they reflect causally, in normal circumstances, the alterations in specific body parts. These representations (1) are the output of nervous receptors responding to several stimuli in our body and (2) integrate the input of the cognitive processes that are for the first time cognitively revealed to the experience subject. Through the action of cognitive processes operating in the nonconceptual representations the subject becomes aware of his phenomenal states. Becoming aware of his own experiences is a cognitive process that subsumes for the first time the phenomenal experience to concepts. Only at this moment, the experiential subject does adopt an epistemic (cognitive) attitude in relation to his phenomenal states. Below this level of awareness – traditionally called introspection – there is the phenomenal consciousness. The phenomenal content would therefore be better described as a nonconceptual representation of physical alterations, ready to integrate the input of processes that have as output intentional states with propositional contents (beliefs and desires). This way, to be the subject of a phenomenal experience would mean to possess certain nonconceptual intentional contents, which might be inputs of certain cognitive processes, that is, nonconceptual representations that could provide different responses.

Now we can have a whole functional description of the mental events. As the consequence, mind and body can no more be thinking as separate entities. Following that perspective, the Self is the network of a lot of processes that involved the performance of distinct functions, some of them that can be described in a mentalist vocabulary. The Self is neither a transcendent unit that controls all the system, nor some specific part of it. It is the process itself. It can also be projected in the world and modified. The informational set of the receptors can be composed with data located inside the limits of our physical body, as well as with data from the outside. Our memory is not only what we can know by introspection or the pieces of information we can have by an exam of our brain or neurological system, but every external expansion of it. Our books, dairies, computers, and our environment in general take an essential part in our set of information.⁹⁴ That can explain why we feel ourselves lost - or having lost some part of us - when we lost a notebook or the information that we have already recorded in our computer. That can explain also the vivid experience that we have when we go to our childhood houses or the loss of mental capacities in old people, when they are moved from the place where they were living for a long time. The Self is neither an entity related to the world, nor one that only represents that. The Self is a network of processes in the world.⁹⁵

⁹⁴ The idea of an externalized Self is developed in: Daniel Dennett, *Kinds of Minds: Toward an Understanding of Consciousness*, Science Master Series, 1996.

⁹⁵ About that see: W. Teed Rockwell, *Neither Brain nor Ghost*, Cambridge, Mass: The MIT Press, 2005.

II

Now I would like to go to the second question: How can that functionalist description of the mental events and, consequently, of the Self fit our ordinary concept of person⁹⁶ and preserve its most peculiar character?

Until here we can see human beings as a socio-biological system that take part, react to and change their environment. Traditionally, human being is defined as a rational being.⁹⁷ What it means in a contemporary and no-metaphysical sense, I intend to explain based on Joseph Raz explanation of rationality.⁹⁸ Following the most significant philosophical tradition, Raz takes rationality as what makes us into persons. In his words

*Our capacities to perceive and understand how things are, and what response is appropriate to them, and our ability to respond appropriately, make us into persons-creatures with the ability to direct their own life in accordance with their appreciation of themselves and their environment (...)*⁹⁹

In the most trivial sense, Raz characterizes rational beings as every creature that possesses the so called ‘capacity-rationality’. That capacity is described as the ability to perceive reasons and to conform to them. The capacity-rationality involves a range of other capacities as perceptual ability, possession of memory, capacity for conceptual thought, capacities to form and to reach decisions and the capacity to control ones movements. Someone who lacks a minimal degree of some of these abilities fails the preconditions of rationality and is neither rational nor irrational. He lacks capacity-rationality. As the other further abilities, the possession of capacity-rationality is a matter of degree. Our rationality can neither be identified with a process of deliberation or decision, nor expressed as a result of one of that acts. It is expressed automatically in the way we function, as reasons can affect ours believes, desires, decisions and actions.¹⁰⁰

With that conception of rationality, Raz can refuse the traditional distinctions between theoretical and practical rationality¹⁰¹ and the more recently distinction

⁹⁶ Here I will use the word “person” to avoid any kind of speciesism, i.e. the idea that we are special simply because we belong to a certain category, the human species.

⁹⁷ To be rational in same philosophical views, as in Kant, Habermas or Scalon, involves the concern with moral justification and/or act according to moral principles. To make that connection, they need a complex theory of Rationality, that we will not discuss here.

⁹⁸ See, Joseph Raz, *Engaging Reason: on the theory of value and action*. Oxford: Oxford University Press, 1999.

⁹⁹ Idem, *ibidem*, 67.

¹⁰⁰ See, idem, *ibidem*, 69.

¹⁰¹ As proposed by Kant.

between substantive and procedural rationality¹⁰². Until here nothing in his concept of rationality is incompatible with our functional description of the Self. Rationality is presented as a capacity or a function developed with a range of others capacities or functions. The question now would be to explain, what we can take as reasons and how they can affect us. I do not intend to discuss Raz's conception of reason. For my purpose, it is enough to understand reason as pieces of information that can improve the ability of the Self to respond appropriately to different inputs. The Self learns how to identify the most important information and react based on them. That is what we can call "to be sensible to reasons". To sum up, we could call a rational being the one that can behave based on the selection and evaluation of information. How complex or developed is that process of evaluation will determine the degree of rationality of the system.

Is that description really enough to distinguish a person? Sometimes we call a person people that cannot anymore, or never could, respond appropriately their environment; people that lack the minimal degree of the mentioned capacities. Other times we attribute some degree of rationality not only to animals but also to computer or their so called "intelligent machines". If the first situation is really the case, so we should say (i) that the rationality-capacity is not necessary to identify a person or (ii) that our ordinary use of the word "person" overtakes the legitimate use of this concept. If the second context is the case, so we need (i) to look for another distinctive characteristic of persons or (ii) to accept that we are just one more rational system in the world. Before deciding for one of these positions, let us look for others characterizations of persons.

We can call an animal or a machine rational, but we do not usually say that they are free or responsible for their action. Why not? Because we think that the human action involves something special; something that can be a sign of the concept of Freedom.

In the article "Freedom of the Will and the Concept of a Person"¹⁰³ Harry Frankfurt defends that the essential distinction between persons and other systems lies on the structure of will. Persons are not the only ones to whom we attributed mental states, that have desires or are capable of taking decisions, but they are the only ones that possess the capacity of constituting a desire of second level.¹⁰⁴

Following that conception, we speak of a desire of the first level when someone tries to achieve what he wants, and we speak of a desire of the second level

¹⁰² See, D. Parfit, "Reason and Motivation". *Aristotelian Society*, suppl. 71 (1997), 99, at 101. Against Parfit's position, see Raz, 72-73.

¹⁰³ Frankfurt, H., "Freedom of the Will and the Concept of a Person", *The Journal of Philosophy* 68 (1971).

¹⁰⁴ See idem, *ibidem*, 6.

when someone would like to possess some desires or would like some desires to constitute his will. In this second case, we speak of volitions of the second level. The essential remark of the concept of person is by Frankfurt provided not only by the presence of desires of the second level in general but by the presence of volitions of the second level. Because a person has volitions of the second level, he is able to enjoy a free will.¹⁰⁵

What does it exactly mean to speak of level of desires and of freedom of will? According to the ordinary concept of freedom, someone is free when he can do what he decides or wants.¹⁰⁶ Against this definition of freedom, Frankfurt presents the concept of freedom of will. To be free, in this way, is not to be free to do what someone wants or decides to do, but to be free to desire what he wants.¹⁰⁷ But what does it mean to say that someone is free to desire what he wants? What is really implied in the concept of freedom of will that lacks in the ordinary concept of freedom?

According to common sense, we can only charge someone for the consequences of his acts when we can assign to him the freedom of acting according to his own decisions. My thesis here is that, to make the agents responsible for his acts, we need neither to imagine levels of desires, nor to introduce the concept of freedom of the will. Freedom as the capacity of determining the own actions is what we can describe in a most trivial sense as the capacity to act based on the evaluation of beliefs and desires.¹⁰⁸

Following that idea, we say that someone is responsible for his acts when we realize that he could have acted in another way if he had decided it. In other words, when we realize that he has not only acted without cohesion but also that he was able to choose among alternatives. The absence of cohesion and the presence of alternatives seem to us as a background to consider someone as free and, therefore, as responsible for his actions. However, we can imagine situations where the agent act,

¹⁰⁵ See *idem*, *ibidem*, 11-14.

¹⁰⁶ See *idem*, *ibidem*, 14.

¹⁰⁷ See *idem*, *ibidem*, 18-19.

¹⁰⁸ That capacity is what we sometimes call deliberation-capacity. In this article, I have intentionally avoided to use the concept of deliberation. The reason for that is because, as I understand it, deliberation is a process, the whole process of checking, selecting, evaluating information and reacting to an action. So if I talk about the several parts of it, I have already talked about deliberation and I don't need to reintroduce it. I know that some authors used to identify deliberation with only one part of that process: the level between the evaluation itself and the answer. They think about that as a decision point that we could discriminate separately. That seems to me as meaningless idea. I just cannot understand what a decision itself can be, without thinking about its contents - i.e. the information and alternatives in case - and about the way it is expressed, i.e. the action. Any other alternative sounds me as a reminiscence of the Cartesians Conception.

free from cohesion, based on his desires, but in spite of that we are not inclined to say that he is responsible for it. These are just the cases when the agents in question are animals, children or adults who lack the minimal abilities to take rational decisions. So we can sign the above mentioned rationality-capacity, i.e. the capacity of act based on the evaluation the most relevant believes and desires as a necessary condition, although not as a sufficient one, for the attribution of responsibility.

The capacity to determine the own actions according to the own aims is what we traditionally call freedom as self-determination. A self-determined being is the one that can act based on reasons: based on the most relevant set of information about his believes, desires and his environment. In most contexts, what characterizes one person and permits us to distinguish him from all other systems can be rescued through the concept of freedom as self-determination. But what that concept of freedom really adds to the description of rational being previously mentioned? From the point of view of the functional description I would say: Nothing. The difference stays in the background. It only makes sense to talk about self-determination when there is not external cohesion, in other words, when the result of the process can be determined internally for her several stages. To guarantee this, authors such as Kant and Frankfurt introduced a conception of a Will disentailed of the natural causal law: A metaphysical Will. If we follow the functional description here presented, it does not seem difficult to comprehend how the several stages of a process can generate autonomously an answer. Using the trivial example of the machine that supplies us a can of coke, we would tell that the process autonomously generated the reaction when we did not kick the machine to withdraw the can or turned off its energy source. In more complex cases, we attribute the self-determination power when the background allows us to exclude factors that would generate a radical unbalance in the weight attributed to the system. Factors of this kind are what we call cohesion, be them external or internal to the corporeal limits of the system. Thus, in the example of typical human actions, our background excludes the attribution of freedom in cases of serious internal disturbances, as well as in cases of external great violence.

But is that description really enough to distinguish people from machines and other animals? My answer is: Yes and No. Yes, if we can accept that what distinguishes us is only a matter of the degree and complexity of some functions. No, if we insist on believing that something substantial can mark the difference of our form of existence in the world. I do not believe that to justify our attitudes as persons we need to look for some other characterization.¹⁰⁹ But if we accept that all the difference is a matter of degree, then we can imagine a new configuration in the

¹⁰⁹ A human being's interesting characterization is supplied by Alasdair MacIntyre in his book *Dependent Rational Animals: Why Human Beings Need the Virtues*. Illinois: Open Court Chicago and La Salle, 1999. MacIntyre focalizes on our basic dependence of the cares of other human beings.

future, and we should be open to the possibility of expanding the same consideration or respect that we demand today for the current "human beings" to other systems. This, however, does not seem to be a problem. We have already enlarged the scope of our respect or consideration (or the scope of those to which we attribute some right), and we also enlarge our moral language to include the respect to others forms of life every day. An example of this is the defense of the third generation rights.

But what we should do with respect to the so called "human beings" that cannot supply that description? In other words, what to do in relation to them, that do not have rationality-capacity or that present just a little degree of the other backgrounds mentioned capacities? This seems to be the most important question of the bioethics discussion. I do not intend to answer it thoroughly, because this would demand another article. But I suggest that some revisions must be made. A functionalist description of the Self has, as a consequence, a demystification of the human and/or person idea.¹¹⁰ In this sense, to admit that some systems are not anymore characterizes as human and/or person is not a sacrilege. From the moral point of view, the most immediate consequence is a decentralization of the human perspective or the end of anthropocentrism. It does not matter whether some creatures are or not human/person. What is important is our commitment to improve the set of capacities of the different systems. So we need to expand our comprehension of the several functions and enlarge our life quality standards.

In Medical¹¹¹ and in political concerns¹¹², Quality of Life Index has been used to guide our mutual conduct. What is the fair treatment in each case will depend on the set of capabilities, i.e. on the functions that which system accomplish.¹¹³ That functions can include very simply ones - as to be nourished and avoiding suffering - as well as the most complex ones, as to be engaged in a political society, to educate children, to paint or to write a book. Based only on a functional description of the objects of our moral concern, we do not have any reason to discriminate some forms

¹¹⁰ Here I intend to refuse the human as the person prejudice, i.e. humanism and personism. About the dispute between humanism and personism see: P. Singer, *Practical Ethics*, Cambridge: Cambridge University Press, 1993 and B. Williams, "The Human Prejudice", in B. Williams and A. W. Moore (eds.) *Philosophy as a Humanistic Discipline*, Princeton: Princeton University Press, 2006.

¹¹¹ An example of that are The Sickness Impact Profile (SIP) developed by Marilyn Bergner and equip, the Quality of Life Index (IQL) developed by Walter O. Spizer and equip and the Health Status Index (HIS) developed by Milton Chen and equip. About that see: Dan Brock, "Quality of Life in Health Care and Medical Ethics", in: M. C. Nussbaum and A. Sen (eds.) *The Quality of Life*, Oxford: Oxford University Press, 1993.

¹¹² That idea is developed specially by Amarty Sen and Martha Nussbaum in several articles. A summary of that discussion can be found in their book *The Quality of Life*.

¹¹³ Here we could also use the terms "capabilities" or "functionings" as Amartya Sen has used in his theory. See: A. Sen, *Inequality Reexamined*. Oxford: Clarendon Press, 1992.

of lives and to suppose that some of them are worthy of respect. But what conduct better expresses our respect will differ from one context to another. To let a system die can express in some cases indifference and to be morally condemnable. In other cases, it can be the expression of how much we care about the suffering of someone, and the recognition, until the last moment, of the eminently human right to the self-determination. It is time to review our convictions, and to make our moral system both more inclusive and more compatible with everything that we have been learning concerning ourselves and our environment.

What Perry doesn't know about Mary

Maria Clara Dias

Wilson Mendonça

Centro de Ética e Filosofia da Mente/UFRJ

Abstract

The paper takes issue with Perry's defense of physicalism against the thread posed by the so called knowledge argument ("Mary's case"). His defense is based on a distinction between the "subject matter content" and the "reflexive content" of knowledge-states. It is argued that, while this distinction can illuminate many cases of newly acquired knowledge, its application to Mary's case presupposes that knowledge of what it is like to have an experience *E* is a case of *factual, propositional* knowledge. In arguing for this presupposition Perry needs to reject the Ability Hypothesis (Lewis and Nemirow), according to which knowledge of what it is like to have an experience *E* is an instance of knowing-how as opposed to knowing-that. The paper claims that Perry's argument for generally subsuming knowing-how under knowing-that fails.

As the legend goes, Mary finally leaves the black-and-white room and sees for the first time, without the intermediary of black-and-white monitors, a red thing, say, a ripe tomato. Mary already knew a lot before she went through the new experience. Indeed, according to the legend, she already had acquired all the knowledge that can be acquired in matters physical. To many spectators of Mary's tale, however, it is plausible to assume that she acquired a new bit of knowledge at the very moment she had the visual experience of the ripe tomato. This new bit of knowledge is usually dubbed "knowledge of what it is like to see red." It gives the physicalist a problem, for its content (the fact that makes Mary's new belief true) is not included in the set of physical facts already within Mary's grasp. Remember that this is supposed to be the *complete* set of *all* physical facts. There would then be at least one non-physical fact; and the claim that physical theory is able to describe at least in principle the whole real world would turn out to be false. Another way of putting the problem is by saying that by having the new experience Mary acquires *phenomenal* information. This has to do with some phenomenal property being instantiated in a certain part of the real world. The phenomenal property cannot be physical, otherwise phenomenal information wouldn't be new to Mary.

Of course it is only under the assumption that "knowledge of what it is like" is a *bona fide* case of propositional knowledge that Mary's tale poses a problem for the physicalist. Under this interpretation Mary's tale becomes a "knowledge argument" against physicalism. Either by denying that Mary's new experience is a case of

knowledge at all or by construing knowledge of what it is like as a species of non-propositional knowledge (for instance as a species of “knowledge how” *as opposed to* “knowledge that”), one can avoid, at least tentatively, the anti-physicalist thrust of the knowledge argument. Under this alternative interpretation, the problem posed by Mary’s tale would be dissolved. Many physicalists, however, accept the initial terms of the problem and try instead to solve it.

John Perry is a case in point. In his recently published book *Knowledge, Possibility and Consciousness*¹¹⁴ he defends what he calls “*prima facie* physicalism” against three famous arguments aiming to show the physical irreducibility of phenomenal consciousness: the zombie argument, the knowledge argument and the modal argument, due respectively to David Chalmers, Frank Jackson and Saul Kripke. (*Prima facie* physicalism—or “antecedent physicalism”—is the claim that we should accept physicalism as true until someone proves the contrary.) Perry’s general claim is that the real targets of these arguments are doctrines that physicalism need not and should not include. Perry tries to isolate in each case the offending doctrine and to show that a suitably conceived physicalism deprived of them, but still responsive to our common sense views about the reality and importance of the phenomenal aspects of experience, remains intact. Concerning the knowledge argument—the only one to be considered here—the offending doctrine is the “subject matter assumption,” a certain view on the content of belief-states underlying knowledge. We will return presently to this point.

Concerning the nature of the state of knowing what it is like, Perry lays his cards quite clearly on the table:

*... when [Mary] exits the room and sees a ripe tomato or a fire hydrant, she will learn something. She will learn what it is like to see red. And this is something new, something she didn't know, and couldn't have predicted, in spite of knowing all the relevant physical facts. She learns a **proposition** about what it is like to have a certain experience, the experience of seeing red. This fact or **proposition** seems to involve a property, the subjective character, that she has never associated with the experience of seeing a fire plug or a tomato or a red thing. She learns an **additional fact** about that experience, a fact that involves this what it is like property. (76) (Our emphasis.)*

Perry’s account of Mary’s new bit of propositional knowledge depends crucially on two assumptions. The first assumption says that simply by having the new experience (and while the new experience lasts) Mary is able to attend to the subjective character of her experience, in a way that is only possible when she is having the experience. Mary is confronted only with her own experience. The first assumption implies, therefore, that Mary uses this experience as an exemplar of a

¹¹⁴ Perry, J. (2001). Numbers in parentheses refer to the pages of this book.

certain type, to which she can point with the “inner demonstrative” *this_i*. As Perry puts it, “when we are attending to a subjective character in the subjective way, and wish to communicate what we are feeling or noticing, we use our flexible demonstrative” (116). The adequate expression of Mary’s newly acquired belief would therefore be, according to Perry: “This_i is what it is like to see red.” *This_i* is a new concept which can only be acquired by someone who is having the experience. Thus the experience is the source of a new concept and *at the same time* the privileged occasion for the application of the new concept to this very same experience.

The second assumption says that, despite being different from the concepts Mary already had, the newly acquired concept picks up the same property referred to by the old concepts. Before leaving the room, Mary had a concept of the subjective character of the experience of seeing red. This concept is Q_R and has been acquired by Mary by reading books on the physico-functional aspects of experiences by normal human beings. This concept points to a physical property. Now Mary is able to pick up this property with the inner demonstrative *this_i*. We have, therefore, two beliefs, one old and the other new. They are expressed respectively by “ Q_R is what it is like to see red” and “ Q_R is *this_i* subjective character.” Their informational contents are, in a relevant sense, identical. In another sense, the pieces of information picked up by these beliefs are different. The second assumption is tantamount to the claim that it would be wrong, when specifying the content of Mary’s new knowledge, to consider only the first sense. Perry’s move here is analogous to a distinction made by Fregeans who associate a certain “fact” to the sentence “Hesperus has a certain property P” and a different “fact” to the sentence “Phosphorus has a certain property P,” in spite of the identity of the referents of “Hesperus” and “Phosphorus.”

From our point of view, the first assumption is more problematic than the second one. Before turning to it, however, we will make some comments on Perry’s justification of the second assumption.

In a relevant sense, Perry’s case against the knowledge argument just is the case for an alternative way of individuating contents of knowledge-states. The usual way of individuation is termed by Perry “the subject matter assumption.” This is the claim that the content of thoughts is to be specified exclusively in terms of the referents of the (non-logical) elements of the thoughts in question: the objects, properties, places and dates referred to by the non-logical elements of thoughts. This notion of content does not provide for a finer-grained distinction, as it is needed to account for Mary’s new belief. Anyway, the insufficiency of this notion of content can be, according to Perry, independently demonstrated by considering other cases of suddenly acquired knowledge. To this effect Perry tells us the stories of Terry, Larry and Gary.

An enthusiast of everything having to do with the American Philosophical Association (APA), Larry lives in isolation at Claremont Hotel and spends his time

reading and memorizing the information about past and future meetings of APA. He simply knows everything about dates, schedules and speakers of APA meetings. In particular, he knows that a Symposium on the philosophy of Mellor is scheduled for April 2, 1999 in the very same hotel where he is living. What he does not know—because he is not allowed to know—is what day it is or even what year he is living in. This information he finally gets when he leaves his room and sees for the first time an up-to-date calendar. The calendar states: Today is April 1, 1999. Larry concludes with satisfaction that the next day there will be a Symposium on Mellor. This has of course all sorts of effects on his behavior. The beliefs expressed by “The Symposium on the Philosophy of Mellor is on April 2, 1999” and “The Symposium on the Philosophy of Mellor is tomorrow” have of course the same subject matter content. But to depict the difference in these beliefs as well as to understand their practical effects on Larry’s behavior, one must take recourse to a different notion of knowledge-content.

Larry’s was the case of someone in possession of all relevant temporal relations, but unable to get an effective orientation until his belief system is anchored in time by an act of identification. Gary, on the other hand, is similarly lost, not in time but in space. He knows all the facts “about the locations of things along Interstate 80—the order of states, cities, towns and villages as one progresses east to west or west to east along Interstate 80” (87). He is actually in one of those locations, in a windowless hut across a gas station called Little America. But that is unknown to him. One day he escapes and comes to learn that “this place is Little America.” He immediately infers that “Salt Lake City is southwest of this place.” Again the novelty of this belief can be accounted for only if in the individuation of its content we go beyond the truth conditions imposed by the belief on its subject matter. For if in the specification of truth conditions we take into consideration only the subject matter, the belief is undistinguishable from the belief expressed by “Salt Lake City is southwest of Little America.” And this belief was already part of Gary’s belief system.

The point of these stories is that real understanding of the nature of the newly acquired information, as well as of its consequences on rational behavior, requires that we go beyond the subject matter content. Larry and Gary already knew “everything” about the “external” referents of their respective intentional states, including the relational properties connecting those referents. What we need in order to describe and explain what is going on when Larry and Gary acquire the new bits of information is another notion of content: “reflexive content.” This is still a matter of truth conditions. But the conditions are imposed *also* on some “internal” elements of the representational vehicle of the beliefs in question. For instance, the reflexive content of the belief expressed by the utterance of

(LA) This place is Little America.

can be specified as follows:

(LA) is true iff the place the speaker of (LA) is attending to and drawing attention to with the use of “this” is Little America.

The fact that the belief’s expression (LA) is mentioned in both sides of this equivalence, as well as the occurrence of “this” on the right side, show that the reflexive content is *not* a subject matter content. The latter may be specified by stating that

(LA) is true iff a certain place along Interstate 80 is Little America.

The next step in Perry’s proposal is the application of the distinction between subject matter content and reflexive content to Mary’s case. This should be the key to undo the trap of the knowledge argument. To this effect, Perry affirms that it is compatible with *prima facie* physicalism to assume that Mary can point to the subjective character of her experience with the inner demonstrative “this_i.” The subjective character is also presumably a physical property, to which Mary was able to “point,” even before leaving the room, with the use of Q_R—or even with an “outer” demonstrative. The beliefs expressed by

- (1) Q_R is what it is like to see red.
- (2) This_i is what it is like to see red.
- (3) Q_R is this_i subjective character.

are indistinguishable if we consider only the subject matter content. The difference made by (2) and (3) can only be captured at the level of reflexive contents. Thus the reflexive contents of (2) and (3) are:

(2) is true iff the act of inner attention to which it is attached is of the subjective character of the experience of seeing red.

(3) is true iff the act of inner attention to which it is attached is of the physical referent of Q_R .

The upshot of all this is that Mary's new knowledge of what it is like to see red is a case of "recognitional" or "identificational" knowledge, as in the previous mentioned stories of Larry and Gary. In Perry's own words:

As in the cases of Larry, Gary and Terry, the change in Mary's beliefs does not result in any new conditions on the truth of her beliefs *given* what they refer to. But it does impose new conditions on the truth of her beliefs, *abstracting* from what they refer to: the condition that the subjective character which is the origin of her old concept is the very one to which she is attending.

That's my account of what Mary learned. (117)

This ingenious solution presupposes, however, the first assumption of Perry's general account, namely that knowledge of what it is like is really a case of propositional knowledge. Perry only provides this purported knowledge with a reflexive content, after having assumed that it *has* a subject matter content. But is the first assumption acceptable? Doubts about it can be suggested by the case of the just invoked Terry. His is also, for Perry, the case of someone acquiring "recognitional" knowledge:

Terry: You know I've really enjoyed talking to you. I'd like to see you again. Can I have your number?

Fran: It's in the book.

Terry: And what's your name?

Fran: It's in the book too. (83)

Even the most complete book is of course of no use to Terry. He would like to know the number of *this* person and, as Perry puts it, "that's a hard piece of information for the phone company to print in the book" (83). Now, if this "piece of information," which would have practical effects on Terry's behaviour, cannot be registered in the book at all, why suppose that acquiring it is an experiential state of propositional character? Terry has to acquire some representation that will afterwards change substantially his belief system and his dispositional attitudes. But for all that, the crucial representation can be non-propositional. From the fact that this

representation can be invoked in the explanation of the later occurrence of propositional attitudes and knowledge-states in Terry one cannot infer that it is essentially propositional—or that acquiring it must be some sort of belief-like attitude.

Thus, Perry's intention notwithstanding, the analogy between Mary and Terry does not lend support to the contention that knowing what it is like to have an experience is a clear case of knowing-that. We agree that in Mary as in Terry an "epistemically relevant change" (125) can be found. What is not obvious is that this change should be seen as the apprehension of "an old fact in a new way," which requires from us the specification of "a new bit of knowledge and a new fact at the level of reflexive content" (125).

With Larry and Gary, the case *seems* to be different. In Larry's utterance "The Symposium on the Philosophy of Mellor is tomorrow" the word "tomorrow" refers to the day identified on the calendar as April 2, 1999. Someone could equally have said to Larry: "The Symposium on the Philosophy of Mellor is on April 2, 1999." Here we have a propositional content as the relevant part of the change in him. It is the same with Gary: "Salt Lake City is southwest of this place" is extensionally equivalent to "Salt Lake City is southwest of Little America" as written on a road map. Here we can plausibly talk of identification, that is, of knowledge in a full-fledged, propositional sense. And the distinction proposed by Perry can be in these cases, which involve the use of spatial and temporal demonstratives, personal pronouns and similar devices, really illuminating. What is at stake here is the generalization of this strategy to the description of all sorts of epistemically relevant changes. The beliefs expressed by "The Symposium on the Philosophy of Mellor is tomorrow" and "Salt Lake City is southwest of this place" occur, as it were, further down the line *after recognition*. What Larry would like to know in the first place is what day *today* is, whereas Gary needs first of all to know the name of *this* place. But these are again hard pieces of information for someone to write in any book. After the acquisition of these pieces of information, that is, after recognition proper takes place, Larry and Gary are able to do all sorts of inferences. The crucial point is that the later occurrence of propositional attitudes as a consequence of recognition proper is no proof that the state leading to them is also propositional. Anyway, refusal to apply Perry's strategy to Mary's case is *not* based on terminological scruples. One can still call Mary's new acquisition "knowledge of what it is like." The important point is that this knowledge is not obviously knowledge-that.

We cannot, therefore, simply assume with Perry that "This_i is what it is like to see red" is an adequate expression of Mary's new state. What is this "this_i" pointing to? What is identified by it? Perry's answer is: a definite type common to many experiences, "a subjective character that can be the character of many experiences of many people" (78). This means that Mary is supposed to be acquiring general knowledge on the way human experiences are, not simply knowledge that something

is now occurring “in her mind” that has never occurred before. Remember, however, that the only source of this general knowledge is her private experience. Mary has nothing else to compare with it. From Perry’s point of view, it is a matter of attending to the experience and recognizing *in it*, in abstraction from all relations to contextual features, the kind of experience that it is. After all, knowledge of what it like to see red is supposed to be something that Mary acquires “in a flash,” without making any sort of relations with other aspects of the world or even of her inner life. The experience must somehow discriminate and identify itself. This makes the what-it-is-like property supposedly identified by Mary a radically intrinsic property. Indeed, Perry states

... that the subjective character of a mental state is not an historical or contextual property of it. It is a property of it that is determined by current inner events. The phenomenal event will typically have external causes and effects, and it may have many current properties that are determined by such external factors. But the subjective character of the event will not be one of these properties. The subjective character is a matter of what it’s like to be in the state, not its typical causes, nor its causes on a given occasion. (35)

Thus, in order to decide exactly which property of her experience the inner demonstrative should be pointed to, Perry’s Mary has no other resources than the particular experience itself. For if she goes *beyond* this experience and undertakes some sort of triangulation of it in a network of external relations, she will be pointing the internal device to the wrong *contextual* property, *not* to the what-it-is-like property. At this point, Perry’s account seems to run afoul of Ludwig Wittgenstein’s private language argument. Indeed, as conceived by Perry, Mary seems to be in the same predicament as the man who tries to fix the meaning of a sensation-word “S” by concentrating his attention on the connection between the sign and the sensation as it first occurs (Wittgenstein 1953, § 258)—or of the man who “peeks” only at his memory images of the timetable in order to make sure that he remembers rightly when the train departs (Wittgenstein 1953, § 265). And if we suppose that Mary’s purported cognitive performance does not have to occur “in a flash” but can be the result of the extended, repeated attending to the lasting experience, then we will fall into the trap of the man who buys many exemplars of the same newspaper to verify the veracity of something he reads there (Wittgenstein 1953, § 265).

If the remarks so far are convincing, knowledge of what it is like is *not* knowledge in a full-fledged, propositional sense; it is *not* know-that. But what is it then? How can we account for the epistemically relevant change in Mary? To be sure, as a consequence of her new experience, all sorts of dynamical relations and abilities emerge in Mary. With the new experience, Mary acquires certain abilities, e.g., the

ability to recognize red things by sight alone, the ability to recognize the experience when it comes again, to remember the experience in question, to imagine a red expanse. If the reasoning above is OK, the explanation of these relations and abilities, which constitute know-how, cannot ultimately appeal to a form of know-that. An alternative explanation is the so called Ability Hypothesis proposed by Laurence Nemirow and David Lewis. “The Ability Hypothesis,” as Lewis (1990, p. 516) puts it, “says that knowing what an experience is like just is the possession of these abilities to remember, imagine, and recognize. ... It isn’t knowing-that. It’s knowing-how.” And according to Nemirow (1990, p. 495), “knowing what an experience is like is the same as knowing how to imagine having the experience.”

The alternative explanation—as in Churchland (1989)—would, therefore, start with non-propositional representations acquired by having experiences. Perry denies this. In his view, Mary’s is only a particular case of a more general dependence relation between know-how and know-that. He definitely regards “knowing how to do something as a species of factual knowledge” (122). “Know-how is a special kind of knowledge of facts about ‘way-of’ relations” (121). This implies that the “way-of” relations—the method by which something can be achieved—are represented in such a way as to be the internal object of a propositional attitude. As Perry puts it: “Know-how is a positive doxastic attitude—i.e. something belief-like, if not paradigmatically belief” (122).

Again, Perry offers no direct support for these bold claims. And again, some of the examples offered by him seem to tell against them. Thus he distinguishes between his wife’s rather “theoretical” knowledge of the movements which are a way of moving the bike in the direction one wants to go without falling and his granddaughter’s “practical” knowledge of how to ride a bike. Involved in both cases is an internal representation “that a certain type of execution will in certain circumstances be a way of bringing about a certain result” (122). As this description blurs the distinction just mentioned, Perry introduces the notion of an *executable* schema/representation, to account for his granddaughter’s abilities, and the notion of an “Humean” idea/representation, as the source of his wife’s knowledge. But then it is not clear that these distinct representations can be interestingly described in the neutral terms of “way-of relations.” Also it is not clear why the access to these very dissimilar representations should be subsumed under the type “belief-like.” It seems more plausible to assume that his granddaughter’s know-how is something inarticulable, something *non-truth-valuable*. Of course, she could *also* acquire articulable knowledge of the correct movements someone should make while riding a bike. This is properly a propositional attitude; and it is an *extra* knowledge-item. The fact that both forms of knowledge can go together, but do not have to go together, is evidence that we are dealing with forms of knowledge independent of each other.

Although Perry states that, in general, the conception of knowledge developed by him “exalts knowing-how” and also that, in particular, he would agree “with

Nemirow and Lewis that Mary's new knowledge is a case of knowing-how" (124), he has a very restricted view of the relation between both forms of knowledge. Thus he affirms confidently that "knowing what it is like to have an experience, and knowing how to do something, are both special cases of knowing-that" (124). It is our concluding contention that Perry's account here simply goes astray. What Mary learns when she enters the color-world is not akin to a propositional content. This is what Perry doesn't know about Mary.

REFERÊNCIAS BIBLIOGRÁFICAS

- ARISTÓTELES. *Metafísica*. Trad. Giovanni Reale. São Paulo: Loyola, 2001.
- BAKER, L.R. *Explaining Attitudes*. Cambridge: Cambridge University Press, 1995.
- BENNETT, J. *Kant's Analytic*. Cambridge University Press: Cambridge, 1966.
- BLACKBURN, S. *Think*. Oxford: Oxford University Press, 1998.
- BRADDON-MITCHELL, D, JACKSON, F. *The Philosophy of Mind and Cognition*. Oxford: Blackwell, 1996.
- BRENTANO, F. *Psychologie vom empirischen Standpunkt*. Leipzig: Duncker & Humblot, 1924.
- CHURCHLAND, P. "Knowing Qualia: A Reply to Jackson." In: Paul M. Churchland. *A Neurocomputational Perspective*. Cambridge: MIT Press, 1989.
- DIAS, M.C., "Kant: os limites da razão crítica", In: *Síntese Nova Fase 72*, Belo Horizonte, 1996.
- DIAS, M.C. *O Argumento da linguagem privada: investigações filosóficas acerca do discurso significativo* (dissertação de mestrado), PUC/RJ: Rio de Janeiro 1989.
- HACKER, P.M. *Insight and Illusion*. Oxford : Clarendon Press, 1986.
- HUSSERL, E. *Logische Untersuchungen*. Halle: M. Niemeyer, 1928.
- JACKSON, F. "What Mary didn't Know". In: *Journal of Philosophy*, 83, 291-295, 1986.
- JACKSON, F. "Epifenomenal Qualia." In: *Philosophical Quarterly*, 32, 127-136, 1982.
- JACKSON, F. *Mind, Method and Conditionals*. London: Routledge and Kegan Paul, 1998.
- KANT, I. *Kritik der praktischen Vernunft*, Berlin: Werke, 1902.
- KANT, I. *Kritik der reinen Vernunft*. Frankfurt: Suhrkamp, 1968.
- KAUFMANN, J. N. "A Problemática da Consciência e a Lacuna Explicativa." Princípios. In: *Revista de Filosofia*, vol. 6, n.7, pp. 87-120, 1999.
- KENNY, A. *Wittgenstein*. London: Penguin Books, 1973.
- LEWIS, D. "What Experience Teaches." In Lycan, W. (ed.) *Mind and Cognition*, pp. 499-519. Oxford: Blackwell, 1990.
- NAGEL, T. "What it is like to be a bat", in Nagel, T. *Mortal Questions*. Cambridge: Cambridge University Press, 1999.

NEMIROW, L. "Physicalism and the Cognitive Role of Acquaintance." In: Lycan, W. (ed.) *Mind and Cognition* pp. 490-499. Oxford: Blackwell, 1990.

PERRY, J. *Knowledge, Possibility and Consciousness*. Cambridge: MIT Press, 2001.

STRAWSON, P.F. *The Private Language Argument*, London: Macmillan St. Martin's Press, 1971.

TUGENDHAT, E. *Einführung in die sprachanalytische Philosophie*, Frankfurt: Suhrkamp, 1976.

TUGENDHAT, E. *Selbstbewusstsein und Selbstbestimmung*, Frankfurt: Suhrkamp 1979.

TYE, M. *Ten Problems of Consciousness*. Cambridge: Bradford Books, 1998.

WITTGENSTEIN, L. *Philosophical Investigations*. Oxford: Blackwell, 1953.