UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

Angela Maria Gomes da Silva

Relações entre Teoria da Identidade e Funcionalismo na Filosofia
da Mente

São Paulo

2011

Angela Maria Gomes da Silva

Relações entre Teoria da Identidade e Funcionalismo na Filosofia da Mente

Dissertação apresentada ao programa de Pós-Graduação em Filosofia do Departamento de Filosofia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, para obtenção do título de Mestre em Filosofia sob a orientação do Prof. Dr. Osvaldo Frota Pessoa Júnior.

São Paulo

2011

Agradecimentos

Gostaria de agradecer à minha mãe e demais familiares que se admiram por eu ainda não ter acabado os meus estudos. À minha família pelo apoio material e pela resignação de eu não prestar concurso público para ganhar muito dinheiro. Meus agradecimentos a Jack Smart por haver me enviado prontamente da Austrália uma cópia atualizada de sua bibliografia sem a qual o meu trabalho teria sido mais árduo e deficitário. Agradeço ao meu orientador por acreditar que eu poderia concluir a dissertação mesmo fazendo um bacharelado na UFABC em Ciência e Tecnologia simultaneamente. Também lhe agradeço as oportunidades de ministrar seminários desde o período de iniciação científica, a liberdade na escolha e desenvolvimento dos temas e também nos feedbacks referentes à didática. Agradeço à Maria Helena e demais funcionárias da secretaria, pela paciência com minha inaptidão com a burocracia da Universidade. Às amigas pelo apoio moral e cobrança pela conclusão da dissertação. Aos amigos e amigas que de alguma forma torceram para que isso acontecesse. Finalmente, agradeço à CAPES.

Resumo

GOMES, Angela. Relações entre teoria da identidade e funcionalismo na filosofia da mente. 2011. 122f. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo.

A teoria da identidade mente-cérebro, que foi desenvolvida no início da segunda metade do século XX, especialmente por filósofos que trabalhavam em universidades australianas, como J. J. C. Smart e U. T. Place, os chamados “materialistas australianos”, sofreu grande impacto de uma teoria que foi desenvolvida logo após, o funcionalismo. A presente dissertação pretende mostrar que não há uma divisão rigorosa entre as duas teorias e que desenvolvimentos da teoria da identidade feitos por David Armstrong e David Lewis, posteriormente considerados como formas de funcionalismo, foram implicações naturais da teoria da identidade.

Palavras-chave: teoria da identidade, funcionalismo, materialismo australiano, filosofia da mente, neutralidade de tópico.

ABSTRACT

GOMES, Angela. Relations between identity theory and functionalism in the philosophy of mind. Thesis (Master Degree) 2011. 122f. – Faculdade de Filosofia, Letras e Ciências Humanas. Departamento de Filosofia, Universidade de São Paulo, São Paulo.

The mind-brain identity theory, developed in the beginning of the second half of the 20th century, especially by philosophers working in Australian universities, such as J. J. C. Smart and U. T. Place, the so called "Australian materialists", suffered a large impact from a theory developed soon after, functionalism. This thesis intends to show that there is not a rigorous division between the two theories, and that the developments done in the identity theory by David Armstrong and David Lewis, later considered as types of functionalism, were natural implications from the identity theory.

Key Words: Identity theory, functionalism, Australian materialism, philosophy of mind, topic neutrality.

SUMÁRIO

1	INTRODUÇÃO	6
1.1	Breve apresentação dos capítulos da dissertação.....	17
2	CONFUSÕES ENTRE O FUNCIONALISMO E A TEORIA DA IDENTIDADE	19
3	FISICALISMO	26
4	IDENTIDADES	33
4.1	Tipos e instâncias (<i>types</i> e <i>tokens</i>).....	33
4.2	Estados, processos e eventos.....	34
4.3	Tipos de identidade.....	36
5	TEORIA DA IDENTIDADE	42
5.1	Dualismo	42
5.1.1	Críticas ao dualismo de feixe.....	45
5.1.2	Críticas ao dualismo de substância.....	47
5.1.3	Críticas a qualquer teoria dualista.....	48
5.1.4	Críticas à teoria de atributos.....	52
5.2	Behaviorismo	55
5.3	Teoria da identidade	60
6	FUNCIONALISMO	74
6.1	Teoria causal da mente	76
6.1.1	Teoria causal da mente, David Armstrong.....	76
6.1.1.1	O papel das disposições.....	80
6.1.2	Teoria causal da mente, David Lewis.....	86
7	NEUTRALIDADE	93
8	HOMEM MÁQUINA E MÚLTIPLA REALIZAÇÃO	97
8.1	Autômatos Cartesianos	97
8.2	Computadores e máquinas de Turing	99
8.3	Aprendizagem de máquina	103
8.4	Homem como mecanismo físico-químico	106
9	RELAÇÕES ENTRE A TEORIA DA IDENTIDADE E O FUNCIONALISMO	111
10	CONCLUSÃO	114
	REFERÊNCIAS	118

1 INTRODUÇÃO

Uma pessoa, que considera a teoria da identidade e o funcionalismo como sendo duas teorias da mente claramente distintas, poderia se surpreender lendo nas obras de J. J. C. Smart argumentos que ora tentam mostrar o quanto elas são próximas, ora o quanto é preciso distinguir uma da outra; ou, então, um livro que na primeira edição foi escrito dentro da teoria da identidade e, na segunda, recebeu um prefácio que o incluía no funcionalismo, sendo que o restante do livro continuou sendo o mesmo, inclusive o título¹: *A Materialist Theory of the Mind*, de David Armstrong (1993). Ou então a constatação de que a partir das ideias de um teórico da identidade, Smart, outros filósofos, Armstrong e Lewis, desenvolveram uma teoria funcionalista que tinha como objetivo garantir a verdade da teoria da identidade, sem que esta nova teoria fosse considerada como pertencente a uma nova linha teórica. Resumindo, o funcionalismo australiano foi uma consequência de ideias presentes na teoria da identidade, especialmente no livro *Philosophy and scientific realism* e no artigo “Sensations and Brain Processes”, de J. J. C. Smart, e o seu desenvolvimento visava defendê-la sem lhe contestar.

Essa surpresa pode acontecer porque é comum que uma pessoa, quando considera a teoria da identidade, só leve em conta um dos três artigos que lhe deram origem na tradição da filosofia analítica, “Is consciousness a brain process” de U. T. Place (1956), “The 'mental' and the 'physical'” de Herbert Feigl (1958) ou “Sensations and brain processes” de J. J. C. Smart (1959). É claro que estas obras e filósofos não são os únicos que alguém associaria à teoria da identidade, poder-se-ia também incluir algum texto de David Lewis, filósofo que compartilhava das mesmas idéias desse grupo, porém ele não é tão associado a essa teoria quanto os anteriores e, na maior parte dos casos, mesmo o artigo de Feigl não tem a mesma visibilidade nos textos básicos de filosofia da mente que os de Place e Smart, embora seja um trabalho bastante inspirador e bem elaborado, tanto em relação à tradição filosófica quanto à análise dos conceitos ‘mente’ e ‘físico’ e de outras características e conceitos relacionados com a mente.

Nessa área, a produção de Place não foi grande e, quanto a Smart, infelizmente a maioria da sua produção em filosofia da mente é composta de artigos, de correspondências trocadas com outros filósofos, que estão disponíveis para consulta na Biblioteca Nacional da Austrália, e, além destes, as suas ideias sobre a teoria da identidade estão integradas em

¹ Não há declaração de que partes do livro foram alteradas, nem com relação a conteúdo filosófico nem com relação a revisão de alguma forma de erro como gramatical, esclarecimento de algum tema, inclusão ou exclusão de trechos ou capítulos.

poucos livros, como *Philosophy and scientific realism* (1963) e *Nosso lugar no universo* (1991), livros que não tratam exclusivamente de filosofia da mente. Isso significa que não há um tratado ou ensaio amplo sobre filosofia da mente ou teoria da identidade escrito por Smart ou Place, um livro dedicado apenas à filosofia da mente. Apesar dessa lacuna, especialmente em Smart, o ponto positivo é a inserção desse tema no plano geral de suas concepções tanto físicas quanto metafísicas, que têm uma acentuada inclinação cientificista.

É bastante provável que esse cientificismo e separação da identidade “translacional” do que se chama de identidade “ontológica”², entre outros fatores como a diminuição de interesse da comunidade filosófica nessa teoria após a crítica de Kripke e o aparecimento do funcionalismo, contribuíram para essa baixa produção. Chamo de “baixa produção” a consideração subjetiva da visibilidade dos trabalhos desenvolvidos em filosofia que se denominam teoria da identidade mente-cérebro. É inegável que ainda hoje há teorias que, de uma forma ou de outra, utilizam argumentos apoiados em uma identidade entre mente e cérebro, mas parece não haver um projeto sistemático de defesa da teoria da identidade desenvolvida no início da segunda metade do século XX. Para apoiar essa afirmação, faz-se necessário distinguir

O cientificismo dos filósofos da teoria da identidade poderia ter contribuído para essa “baixa produção” na medida em que a teoria da identidade não buscava saber qual é a verdadeira natureza dos estados e processos mentais enquanto identificados a estados e processos cerebrais, mas ficaria restrita aos problemas lógicos e filosóficos dessa identificação. Nenhum desses filósofos procurou entrar em detalhes sobre essa natureza, na realidade, sustentam que a identidade da mente com o cérebro é uma hipótese científica plausível e que é tarefa da ciência descobrir essa natureza. Dessa forma, restringiu-se o âmbito da filosofia, demarcando-se nitidamente o papel filosófico e o papel científico. Esses filósofos não desenvolveram uma filosofia da neurociência, não questionam conhecimentos científicos específicos. Além disso, quando se separa identidade translacional de identidade ontológica, pois Smart afirma em (1991, p.161) que o seu “físicalismo é *ontológico*, não *translacional*”, acaba-se se abstendo de desenvolver um rol de traduções, no caso da identidade translacional, de termos ou sentenças sobre a mente para termos e sentenças sobre o cérebro ou processos e estados cerebrais. Neste caso, haveria necessidade de acompanhar os desenvolvimentos científicos para efetuar as traduções de uma em termos da outra. Essas identidades têm lógicas diferentes e a identidade ontológica, que é a adotada pela teoria da

2 Os conceitos de identidade translacional e identidade ontológica são abordados no capítulo 4 desta dissertação, especificamente na seção 4.3.

identidade, é apoiada no reducionismo físico, ou seja, sua ontologia é a ontologia defendida e fundamentada pela física.

No caso de Smart, a tese da identidade entre estados mentais e estados cerebrais era, tanto para ele como para Place e Armstrong, passível de confirmação empírica pela ciência como qualquer outra hipótese científica pura. Portanto, o papel do filósofo nessa área seria um pouco limitado pelo do cientista, deixando-se para este o papel de provar, na área científica, a identidade que eles defendiam filosoficamente. As últimas frases de Armstrong no livro *A Materialist Theory of the Mind*, página 366, são estas: “Uma teoria fisicalista da mente é um mero prolegômeno para uma metafísica fisicalista. Uma tal metafísica, como a teoria da mente, será sem dúvida o produto da junção da investigação científica e da reflexão filosófica.”

Quanto a David Armstrong, é bom ressaltar que ele desenvolve um trabalho de análise conceitual de estados mentais de considerável extensão, e por isso sua inclusão nessa lista junto a Place e Smart diz respeito a não ter desenvolvido, assim como os outros dois, uma filosofia neural ou uma tentativa de explicação da mente em termos dos processos neurais particulares, como encontramos nos filósofos Paul e Patrícia Churchland, por exemplo. Assim, não se encontra nos trabalhos de Place, Smart e Armstrong³ uma descrição dos estados e processos mentais em termos de estados e processos cerebrais: o que há é uma defesa de uma hipótese científica, “estados e processos mentais são iguais a estados e processos cerebrais”, e uma tentativa de fundamentar e desenvolvê-la dentro da filosofia.

Em todo caso, *a falta de interesse filosófico pela teoria da identidade após o surgimento do funcionalismo* parece ser uma explicação mais plausível para o que aqui se chama de “baixa produção” da teoria da identidade. Um dos objetivos deste trabalho é problematizar essa afirmação, ou seja, problematizar a separação encontrada nos livros de filosofia da mente entre a teoria da identidade e o funcionalismo.

Embora a teoria da identidade e o funcionalismo sejam definidos em capítulos posteriores, é bom salientar neste ponto que o que aqui se considera como teoria da identidade não pode ser confundido com outras teorias materialistas, naturalistas ou fisicalistas da mente. Além disso, o objeto de estudo neste trabalho se restringirá na maior parte dos tópicos ao que se pode chamar de “materialistas australianos”, nome empregado por Alan Reeves e citado por Ned Block no texto “Introduction: what is functionalism?”, de 1980, para se referir especificamente aos filósofos David Armstrong, J. J. C. Smart, U. T. Place e David Lewis,⁴

³ O mesmo podemos afirmar para David Lewis.

⁴ Apesar da denominação, apenas David Armstrong é realmente australiano, os outros desenvolveram essa teoria enquanto trabalhavam em universidades australianas no final dos anos cinquenta e início dos anos 60,

tendo, então, uma história bem concisa e delimitada no campo da filosofia da mente. Então, sobre a teoria da identidade, podemos dizer que *surgiu com os artigos acima expressos, causou um impacto inicial positivo, mas perdeu espaço para o funcionalismo e novas formulações do materialismo*. Assim, Smart (1991, p. 167) dirá: “Existe no momento uma filosofia popular da mente que se chama ‘Funcionalismo’; muita gente acha que já suplantou a teoria da identidade mente-cérebro (e obviamente outras filosofias da mente, como o dualismo de corpo e espírito).” Como um exemplo das novas teorias da mente, que se aproximam da teoria da identidade, tem-se o materialismo defendido por John Heil (2003), que se assemelha àquele apresentado posteriormente em “Sensations and Brain Processes” por Smart, e especialmente em *Philosophy and scientific realism*.

Um dos críticos mais bem sucedidos da teoria da identidade foi, certamente, Saul Kripke. Podemos encontrar seu argumento contra a teoria da identidade em *Naming and Necessity*, seu livro de 1972. Kripke argumentará, baseado em “designadores rígidos”, contra a validade lógica de uma identidade mente e cérebro contingente.

Apesar de no cenário filosófico a teoria da identidade entre mente e cérebro ser pontualmente posta em 1956, com U. T. Place, tanto antes quanto depois desse período há um uso do termo “cérebro” para se referir a processos quer intelectuais quer mentais na literatura e na ciência. Assim, Einstein (1934), em seu artigo “Principles of Research”, falando sobre dois tipos de cientistas que há na ciência, dirá: “muitos outros serão encontrados no templo que ofereceram os produtos de seus cérebros neste altar para propósitos puramente utilitários”⁵. Neste exemplo, vemos como o termo cérebro (*brain*, em inglês) era usado por uma pessoa do meio científico, mas também é possível listar outros casos na literatura em períodos anteriores ao do desenvolvimento da teoria da identidade que é considerada neste trabalho.

É provável que a proximidade entre as duas teorias (se for assumido que Smart está certo e não há oposição rigorosa entre elas) tenha possibilitado uma migração e aderência de alguns filósofos da teoria da identidade ao funcionalismo. O caso de Armstrong e Lewis

ou então por causa da afinidade teórica. Daí o nome de “materialistas australianos”. Smart ainda hoje é professor da Universidade Monash na Austrália. No caso de David Lewis, que não teve o mesmo vínculos de trabalho que Place e Smart, lê-se, na introdução ao livro (Lewis, 1999, p. 5), que Lewis denomina-se ‘materialista australiano’: “I am an ‘Australian materialist’”. Também Sergio Moravia (1995) utiliza o termo *escola australiana* para se referir a Place, Smart e Armstrong como sendo figuras líderes dessa escola.

5 “many others are to be found in the temple who have offered the products of their brains on this altar for purely utilitarian purposes”. Embora os artigos presentes no livro de onde a citação foi retirada, *Essays in Science*, não apresentem as datas originais nas quais foram escritos, o livro do qual esses artigos formam retirados, *The world as I see it*, é apontado no livro *Escritos de maturidade: artigos sobre ciência, educação, religião, relações sociais, racismo, ciências sociais e religião* como contendo materiais que vão de 1922 a 1934.

representam uma indicação dessa proximidade. Um dos objetivos desta dissertação é justamente pesquisar os elementos que aproximam as duas teorias e não se considera a força que outras posições filosóficas desempenharam no desenvolvimento da teoria da identidade e do funcionalismo.⁶ Mas aparentemente as duas são comumente distintas, embora seja corrente que o funcionalismo suporta pelo menos um tipo de identidade, aquela de instâncias⁷, ou seja, que instâncias de estados mentais são identificadas a instâncias de estados cerebrais. Por exemplo, ‘a dor que José sente no braço às 22 horas e 36 minutos do dia 01 de novembro de 2010 é igual ao estado no cérebro de José às 22 horas e 36 minutos do dia 01 de novembro de 2010’. Será mostrado no tópico sobre tipos e instâncias que a forma como se define a palavra tipo também permitirá que essa identidade não apresente muitos problemas.

Voltando ao assunto, apontar que as duas teorias são próximas não é suficiente, importa verificar se elas são equivalentes ou se uma delas apresenta mais vantagens com relação à outra. Apesar disso, questões de valor, equivalência e eficiência das teorias não serão objetos deste estudo, no qual serão focalizadas as características que as aproximam. As duas teorias fazem uso de uma linguagem neutra na descrição de estados mentais, embora, se for examinado com mais rigor, o funcionalismo seja mais inclinado a fazer definições de conceitos enquanto a teoria da identidade se empenha em fazer descrições ou relatos de experiência, especialmente até os estudos conceituais de David Armstrong. Uma ilustração disso é o fato de que no artigo “Sensations and Brain Processes”, Smart estará, como em textos posteriores, preocupado com os relatos de experiência, enquanto isso, a segunda parte do livro *A Materialist Theory of the Mind*, de Armstrong, que é aquela na qual a *teoria do estado central* é desenvolvido ou aplicada, diz respeito justamente à análise e à definição de conceitos mentais como ‘desejo’ e ‘intenção’.

Pode parecer meio equivocado distinguir entre conceitos mentais e experiências, já que supostamente o conceito de dor deve envolver uma experiência de dor, por exemplo. Talvez uma classificação, que pode ser encontrada em Ned Block, no texto “Introduction: What is functionalism?”, ajude a visualizar essa diferença de papel das duas teorias, quando ele faz uma distinção entre dois tipos de funcionalismo: a *especificação funcional*, representada por Armstrong e Lewis, e a *identidade de estado funcional*, representada por

6 Neste caso, está-se trabalhando com apenas duas teorias: funcionalismo e teoria da identidade. Existem outras teorias da mente como: monismo anômalo, formas de dualismo, emergentismo, fenomenalismo, etc. que não estão sendo consideradas. Além disso, não há quantificação de quantos filósofos apoiavam uma teoria A em um momento x e quantos filósofos abandonaram a teoria A em um momento x’ para ingressar em uma teoria B. Não foi feito estudo quantitativo, embora essa classificação seja corrente no meio filosófico. A teoria da identidade e o funcionalismo são consideradas, nesse meio, como sendo duas teorias predominantes em momentos históricos diferentes, sendo a teoria da identidade anterior ao funcionalismo.

7 Neste trabalho, o termo em inglês “*token*” será traduzido por “instância”.

Putnam e Fodor. A identidade de estado funcional define um estado mental através de sua identificação com um estado funcional e então atribui esse estado à pessoa como uma propriedade, resultando que estados mentais são idênticos a estados funcionais e não a estados físicos. Assim, ela se constitui por um processo com dois momentos, um momento de identidade de estados e um momento de atribuição de propriedades. Um exemplo corriqueiro desse tipo de funcionalismo é a definição de carburador, que seria definido apenas por meio de suas relações funcionais sem fazer uso dos componentes materiais de um carburador específico. Por exemplo, carburador é o que mistura ar e combustível em um motor; à peça em um motor que desempenha esse papel, atribui-se a propriedade carburífica ao objeto, mas não se identifica. Já a especificidade funcional define um estado mental como um estado que é especificado funcionalmente, e levanta a possibilidade de esse estado assim especificado ser um estado cerebral. Por hipótese, haveria duas identidades ou descrições nesta segunda forma de funcionalismo. Como na teoria anterior, haveria um momento no qual se identifica o estado mental com um estado funcional e posteriormente se identifica esse estado com um estado físico.

Essa distinção é apresentada de forma um pouco mais simples em Heil (2004, pp. 99-100), quando ele afirma que Armstrong e Lewis identificam os estados mentais a estados funcionais, mas identificam ainda estes aos realizadores, ou seja, ao *ocupante* de um papel causal particular. Dizendo de outra forma, tomando com uma certa liberdade a afirmação de Heil, estes funcionalistas procuram no mundo a coisa que desempenha esse papel causal ou a instância desse estado assim especificado. Já a outra linha funcionalista identifica os estados mentais apenas com os *papéis*, ou seja, a descrição relacional abstrata.

Pode-se dizer, então, que esse ponto de vista de estado mental como sendo uma coisa real, com possibilidade empírica de ser um estado cerebral, é uma característica tanto da teoria da identidade quanto do funcionalismo que surgiu a partir dela⁸. É uma possibilidade porque os teóricos da identidade parecem abdicar de provar a identidade entre mente e cérebro através da metafísica/filosofia pois, como foi dito acima, trata-se de uma hipótese científica e como tal deve ser confirmada empiricamente pela ciência. Não é a filosofia que dará a palavra final sobre a identidade dos estados mentais e cerebrais. Essa posição é corroborada também pela seguinte passagem do capítulo 6, “The Central-State Theory”, da primeira parte do livro *A Materialist Theory of the Mind*, intitulada “Theories of Mind”:

8 Esta minha consideração da “a teoria da identidade e do funcionalismo que surgiu a partir dela” é ambígua e, como já informado, será problematizada. Neste ponto, aceito a classificação de Block e do próprio Armstrong ao considerar a sua filosofia e a de David Lewis como sendo teorias funcionalistas.

A natureza intrínseca desses efeitos e causas não é algo que está envolvido no conceito de mente ou dos conceitos mentais particulares. [...] O que ele é em sua própria natureza é algo para ser descoberto pela ciência. A ciência moderna declara que esse mediador entre estímulo e resposta é de fato o sistema nervoso central, ou, mais crua e inacuradamente, mas mais simplesmente, o cérebro. (ARMSTRONG, 1993, p. 79)

Nesta primeira parte do livro, ele apresenta algumas teorias da mente. A parte 2 desse livro trata especificamente do conceito de mente. E na página 356, ele dirá: “Agora estamos nos movendo no reino da ciência e fato empírico, não no reino da possibilidade lógica que é apropriado para análise conceitual.”

Nesse cenário, o conceito de experiência é muito relevante. É na experiência subjetiva que surge a evidência de que existe alguma coisa além de comportamentos e disposições. Ela torna concretos os estados mentais dos quais estas teorias tratam. Não é à toa que um dos artigos de Smart, no qual nos deteremos mais adiante, é intitulado “Reports of immediate experiences” (1971). Isso também já estava claro no artigo “Sensations and brain processes”, quando ele responde à quarta objeção dizendo que não está afirmando “que a pós-imagem é um processo cerebral, mas que a experiência de ter uma pós-imagem é um processo cerebral. É a experiência que é relatada no relato introspectivo”. Também para outros filósofos, por exemplo, Galen Strawson em *Mental Reality* (1994, p. 118), a experiência é um termo que diz respeito aos acontecimentos mentais (*mental goings-on*), a tudo que de alguma forma é consciente. Place e Feigl também trabalham com objetos mentais próximos desse conceito, todos relacionados a sensações, consciência.

Desse modo, mesmo que tanto os filósofos da teoria da identidade quanto os do funcionalismo causal apontem a localização dos estados mentais na realidade física, há uma diferença, neste caso, entre as duas abordagens. O sentido de estado e processo mental no qual Smart está preocupado diz respeito ao que *está acontecendo* “agora”, sua experiência em um certo momento. Por outro lado, tanto Armstrong quanto Lewis estão mais próximos de uma busca por definir *objetivamente* conceitos mentais. Smart se contrapunha ao *behaviorismo* de Gilbert Ryle para defender a existência dessas experiências internas como *entidades reais*. Já os outros dois filósofos tinham nas mãos outras questões, mesmo que também estivessem preocupados em definir os estados mentais como *entidades reais*. Em Armstrong, isso fica claro quando ele faz uma distinção entre sua teoria e o behaviorismo, pois ele define um estado mental como “*um estado da pessoa apto a ...*” [itálicos meus], alegando que o

behaviorismo define um estado mental como uma disposição e não algo que tem uma disposição.

Entre esses conceitos e experiências, embora seja compreensível que uma teoria limite o mental às coisas que são conscientes, quer atualmente quer em um momento determinado, muitos processos e estados podem ser chamados de mentais e ainda assim não serem nem virem a ser conscientes. Se alguns estados mentais só se tornam conscientes após reflexão e análise, então é provável que um grande número de estados mentais nunca venha a ser conscientes, pois sequer houve ou haverá reflexão sobre todos eles. Feigl fará uma análise dos estados mentais considerando também os estados inconscientes da mente.

Algumas vezes, uma forma de tratar o funcionalismo é compará-lo à engenharia reversa. Tem-se uma caixa preta e a partir do modo como ela responde às várias entradas e saídas de dados, tenta-se descrever seus estados internos. Esse método não é infalível, claro, dependendo do engenheiro pode haver mais ou menos estados nessa descrição do que na máquina original, ou seja, ele pode descrever o interior da máquina com maior ou menor acerto. Mesmo abrindo a máquina para saber como e quais peças estão relacionadas entre si, ele pode não saber a real funcionalidade de uma das peças. Por isso ele não pode garantir a verdadeira forma como os estados mentais estão relacionados entre si e com estímulos e comportamentos. E o mesmo pode-se aplicar para o funcionalismo na teoria da mente. Haverá a oportunidade de acompanhar como Armstrong define os conceitos mentais, como e quais relações ele estabelece para os conceitos. Pois, para determinar uma fórmula para definir os estados mentais e identificá-los contingentemente com os estados cerebrais, ele utilizará um ponto de vista da psicologia sobre o homem, na qual ele é considerado como uma máquina na qual alguns estímulos produzem algumas respostas, sendo que os estados mentais seriam posicionados na cadeia causal entre os estímulos e as respostas. Ao atribuir, dessa forma, estados a uma pessoa, pode-se atribuir também estados mentais inconscientes se eles forem mais propícios a uma melhor explicação.

De qualquer forma, seria muita presunção considerar que todos os estados mentais são conscientes, principalmente se se considera a identidade entre estados mentais e cerebrais, por isso há uma margem para esse tipo de estado mental. Armstrong considerou esse tipo de estado. Para ilustrar, no texto “The nature of mind” (1980), é apresentado um exemplo de um motorista que dirige por uma longa distância sem estar consciente sobre o que está fazendo ou mesmo sem saber se estava consciente de outras coisas. Ele prossegue afirmando que, considerando esse caso, “é óbvio que em algum sentido processos mentais ainda estão ocorrendo quando alguém está num tal estado automático” (Armstrong, 1980, p. 197). Com

isso, apesar da importância da consciência como tema filosófico, não há razão para desconsiderar a possibilidade da existência de estados mentais inconscientes, quer eles sejam considerados como experiência, o que acho plausível, quer não. Em que sentido afirma-se serem os estados mentais inconscientes uma experiência?

Quando se fala em experiência, há quase como pressuposto a existência de um sujeito, aquele que experiencia (tratando-se de uma ocorrência subjetiva, um fenômeno interno ao sujeito que experiencia) e/ou que conduz a experiência (um experimento científico pode ser um exemplo desse segundo tipo, como também um sujeito que conduz, aplica, um procedimento experimental; assim, experiência poderia ser um dado arranjo e procedimento sobre esse arranjo). Neste caso, tratamos no primeiro sentido, porque o segundo pode envolver a criação e implementação de uma experiência física no mundo exterior e aqui interessa a experiência subjetiva, ou seja, aquele que experiencia suas sensações, percepções, pensamentos, etc. Sendo assim, um estado mental inconsciente poderia ser aquele que não é experimentado em um determinado momento, mas que está acontecendo a despeito disso. Se se define experiência subjetivamente, então corre-se o risco de ser considerado como experiência apenas o que o sujeito está experimentando (conscientemente). Um problema similar é discutido por Armstrong e Norman Malcolm no livro *Consciousness & Causality*. O caso é o de um homem com dor de cabeça. Durante uma caminhada ele esquece a dor de cabeça, mas quando alguém pergunta se ele está com dor de cabeça, ele afirma que está. Então, a pergunta é se durante o momento em que ele está prestando atenção a outra coisa e não sente a dor, não está consciente dela, pode-se dizer que ele ainda está com dor de cabeça? A forma como Armstrong define o termo “estado” permitirá que haja inclusão de estados inconscientes entre os estados mentais, mesmo que durante um período de tempo P , em intervalos menores de tempo p dentro de P , a pessoa não tenha sentido conscientemente a sua dor. Dando um exemplo mais concreto, pode-se considerar que uma pessoa teve uma semana de ansiedade, mesmo que ela não tenha estado ansiosa todos os dias ou algumas partes do dia, como durante as manhãs porque estava muito ocupada em seu trabalho.

Retomando à especificidade funcional⁹, a forma de funcionalismo de Armstrong e Lewis apresentada acima, talvez se torne mais compreensível um outro aspecto que a teoria da

9 Apesar de utilizar nestas passagens da introdução o termo “especificidade funcional”, e de que, dependendo do tempo, da forma de análise e citação da palavra, outros termos poderão ser utilizados se forem considerados como sinônimos. Por exemplo, a utilização do termo “teoria do estado central” seria mais adequada por indicar o desenvolvimento da teoria mais contíguo temporalmente. Mas, no corpo do trabalho, será utilizado quase que exclusivamente o nome de “teoria de estado central”, “materialismo de estado central” ou ainda “teoria causal da mente”. Os nomes mudarão de acordo com o texto e passagem considerados. O uso do nome “especificidade funcional” tem apenas uma intenção didática e está relacionado com a análise de Block e sua influência no desenvolvimento deste trabalho.

identidade compartilha com o funcionalismo, a possibilidade da múltipla realização do mental. À primeira vista seria natural ser incompreendida a afirmação de múltipla realização na teoria da identidade porque este aspecto é utilizado normalmente como argumento contra essa teoria e associada ao funcionalismo como uma espécie de diferencial (em alguns casos), principalmente à identidade funcional (funcionalismo computacional). Com efeito, sem conhecer outros textos de Smart, principalmente o livro *Philosophy and Scientific Realism*, não se poderia tomar conhecimento dessa possibilidade teórica.¹⁰ Como expresso acima, a especificidade funcional especifica funcionalmente estados mentais que possivelmente são cerebrais, sendo estes estados as coisas reais a partir das quais se obtêm as funções.¹¹

Para adiantar um pouco o tema da multiplicidade de realização do mental na teoria da identidade e apontar esquemática e brevemente os pontos da teoria que sustentam essa possibilidade, faz-se útil tomar conhecimento de que Smart faz uso de um mecanicismo que aparece desde “Perceptions and Brain Processes”, e que recebe um melhor desenvolvimento no livro *Philosophy and Scientific Realism* (1963).

Primeira consideração (mais evidente): Segundo o ponto de vista mecanicista de Smart, o homem nada mais é que uma máquina muito complicada, tendo como base um mundo no qual todas as coisas que existem são físicas. Além disso, os estados mentais seriam identificados com processos cerebrais em nível neuronal, podendo-se fazer a identificação de estados mentais a padrões de redes de neurônios. Esses padrões poderiam, então, ser implementados em outras “máquinas”, levantando a possibilidade de inteligência artificial. Ao contrário do que possa parecer, a inteligência artificial corroboraria a teoria da identidade, e não a refutaria, pois ela seria a prova de que o ser humano é um mecanismo físico. Na maioria das vezes, esses filósofos falam apenas de processos físico-químicos no cérebro e com isso querem dizer “mecanismo físico” complicado. Assim, se uma máquina, que é um mecanismo físico, pode desenvolver e apresentar mentalidade, então o homem, também pode ser considerado um mecanismo físico dotado de mentalidade, não seria necessário postular entidades não-físicas para explicar a mente, confirmando, com isso, a teoria da identidade. E isso se estende não apenas para máquinas, mas também para outras formas de vida como animal, extra-terrestre (se existir) etc.

10 Embora não se questione neste momento se Smart excedeu a teoria da identidade e seria melhor classificado nesse livro como um funcionalista ou se a teoria da identidade suporta uma identidade de múltipla realização.

11 Os materialistas consideram os estados mentais reais por serem estados físicos, utilizando uma identidade ontológica, ou seja, a cada estado mental corresponde um estado cerebral, e ambos são uma e mesma coisa. Assim, mesmo considerando apenas estados mentais em suas análises conceituais, Armstrong estaria na verdade considerando estes estados enquanto seriam identificados a estados cerebrais. Portanto, a experiência mental que é objeto do relato nada mais é que um estado físico, e por isso real.

Não fica claro se é uma vantagem para esses filósofos, mas partir de padrões neurais facilita a constatação, de forma mais concreta, de semelhanças e diferenças, quer entre os estados mentais de uma mesma pessoa, quer comparando-se seus estados mentais com os de outras pessoas ou de outras espécies, sejam elas naturais ou artificiais, através da comparação das semelhanças e diferenças entre as suas estruturas neurais. A desvantagem pode ser a negligência de outros aspectos que estão ausentes no nível dos neurônios mas que podem ser bem sucedidos na modelagem da inteligência ou mentalidade.

Segunda consideração (menos evidente): Estes filósofos desenvolvem uma teoria da identidade mente-cérebro que não afirma simplesmente que a mente é igual ao cérebro. Eles defendem uma identidade contingente e, por isso, desenvolvem algumas fórmulas “tópico neutras”, ou seja, que seja neutra entre o materialismo e o dualismo, para possibilitar que essa identidade possa ser expressa. Isso fica muito evidente em Smart, que desenvolveu uma fórmula tópico neutra para responder à terceira objeção à teoria da identidade no artigo “Sensations and Brain Processes”. Essa objeção dizia que mesmo se se aceitasse a identidade entre os processos mentais e cerebrais, ainda haveriam *propriedades* mentais que seriam irreduzíveis. Smart, então, desenvolve uma fórmula na qual os relatos mentais são feitos sem que haja afirmações sobre a natureza do que está sendo relatado. Assim, um relato de uma sensação teria a forma “o que está acontecendo em mim é semelhante ao que acontece em mim quando ...”, sendo os ‘...’ substituídos por uma descrição exterior de uma experiência. Esta é a forma geral da fórmula de Smart, que varia um pouco quanto ao que substitui os “...”.

As fórmulas de Armstrong foram desenvolvidas para possibilitar a identificação contingente entre os estados, processos e eventos mentais e cerebrais. Ele reconhece sua *ligação* com a fórmula de Smart, e considera sua fórmula como sendo mais elaborada e que explicaria melhor essa identidade. Embora essas fórmulas sejam consideradas em um capítulo mais apropriado, capítulo 7 desta dissertação, ele a expressa da seguinte forma (fórmula principal): “estado da pessoa apto a causar certo gênero de comportamento” e, para alguns estados, também na forma (fórmula secundária): “estado da pessoa apto a ser causado por certo gênero de estímulo”.

Lewis também desenvolverá uma forma neutra para identificação entre estados mentais e cerebrais, mas sua formulação será mais lógica e abstrata. No artigo “Psychophysical and Theoretical Identifications”, ele apresenta uma formulação para explicar os conceitos mentais a partir de uma *sentença de Ramsey*. Ele também apresenta sua forma de considerar os estados mentais, semelhante à de Armstrong:

[...] um estado mental M (quer dizer, uma experiência) é definível como o ocupante de um certo papel causal R – que é, como o estado, de qualquer gênero [*sort*], que é causalmente conectado em modos específicos a estímulos sensoriais, respostas motoras e outros estados mentais. (LEWIS, 1980, pp. 207-208).

Apesar disso, a possibilidade de múltipla realização pode trazer alguns problemas para uma teoria que trata de identidade. Como foi mencionado acima, Kripke levantou uma forte objeção para a identidade contingente sustentada por esses filósofos. Mas Kripke é essencialista, como sua concepção de ciência também o é, e isso contrasta com a filosofia da ciência dos filósofos que são abordados aqui. Escolher entre uma concepção ou outra de epistemologia da ciência pode enviesar a adesão a uma ou a outra teoria, ou até a nenhuma delas, e a escolha de uma entre tantas teorias da mente.

1.1 Breve apresentação dos capítulos da dissertação

A dissertação se desenvolve em torno dos elementos da teoria da identidade que a aproximam do funcionalismo, aspectos da teoria que possibilitaram o desenvolvimento da “teoria causal da mente” por David Armstrong e David Lewis.

Um dos principais capítulos da dissertação é o segundo, no qual são apresentadas algumas evidências textuais das *confusões entre a teoria da identidade e o funcionalismo*, por isso ele vem antes dos capítulos dedicados a explicar cada uma das teorias isoladamente. Essas evidências são afirmações feitas por filósofos como Smart, Armstrong e Block. A ordem na qual o capítulo é inserido no texto é proposital e visa justamente evidenciar e reforçar a relação ora apresentada.

Os capítulos 3 e 4 são dedicados à apresentação dos conceitos que ajudarão a entender melhor as teorias consideradas, especialmente a teoria da identidade. A seção 8.2 poderia fazer parte do capítulo 4, já que apresenta conceitos que são essenciais para a compreensão do funcionalismo. No entanto, preferi deixá-lo no capítulo 8 porque ele representa uma evolução no conceito de máquina e, com isso, tem implicações na definição de “homem máquina”, conceito que ajuda a problematizar a “múltipla realização” na teoria da identidade.

O terceiro capítulo apresenta o conceito de fisicalismo, introduzindo as principais visões dos filósofos que são estudados nesta dissertação. Escolheu-se abordar os pontos de vista mais cientificistas desses filósofos, quer dizer, não problematizando as implicações dos

posicionamentos metafísicos por eles defendidos. Isso significa que o capítulo é bastante expositivo e não envolve aprofundamento crítico em problemas metafísicos.

O capítulo quarto pretende fazer uma exposição dos principais tipos de identidade, com vista a ajudar o leitor da dissertação sobre os conceitos que estão envolvidos nas teses principais das teorias em consideração. Este e o capítulo anterior são capítulos preparatórios para as duas teorias a serem discutidas. Além disso, também são apresentadas as definições de alguns conceitos importantes, como estado, processo e evento.

No quinto capítulo, será abordada a *teoria da identidade*. Antes disso, também serão feitas breves considerações sobre duas outras teorias da mente, o dualismo e o behaviorismo, em razão da relação existente entre elas.

No sexto capítulo, é apresentada uma visão mais geral do funcionalismo, apresentando principalmente as teorias desenvolvidas pelos funcionalistas computacionais, seguida da análise das teorias causais de David Armstrong e David Lewis, ambos considerados como funcionalistas, seguindo a classificação de Ned Block.

No sétimo capítulo, será analisada a neutralidade de tópico, quando serão consideradas as formas tópico neutras desenvolvidas por estes filósofos.

O capítulo seguinte, oitavo, apresenta a forma como os conceitos de *homem máquina e múltipla realização* estão relacionados com as teorias abordadas. Na verdade, ele pretende proporcionar o conhecimento histórico e conceitual que está subjacente às ideias dos teóricos abordados. Considero que seja um pouco como uma espécie de apêndice, talvez ele consiga mostrar como um teórico da identidade mente-corpo cientificista, como Smart, não pôde abdicar de incluir em sua teoria máquinas de Turing e considerações sobre inteligência artificial, embora esta análise não leve a qualquer consideração sobre seu sucesso nesse intento.

No nono capítulo, é apresentada a relação entre o funcionalismo e a teoria da identidade. Será interessante porque trata da problematização da distinção entre as duas teorias apontada acima. Este capítulo é uma espécie de extensão do segundo, tratando especialmente das *relações entre teoria causal e teoria da identidade*.

Finalmente, o último capítulo apresenta uma breve conclusão sobre o assunto desenvolvido.

2 CONFUSÕES ENTRE O FUNCIONALISMO E A TEORIA DA IDENTIDADE

Neste capítulo, serão mostradas algumas evidências textuais da confusão de classificação de filósofos em funcionalistas ou teóricos da identidade, além da indicação das semelhanças entre as duas teorias. Como o filósofo mais controverso é J. J. C. Smart, ele será quem receberá mais atenção.

Em 1995b, J. J. C. Smart escreveu o “Postscript (1995)” ao seu artigo “Sensations and Brain Processes” (Smart, 1959), confirmando as ideias que foram apresentadas neste artigo e afirmando, entre outras coisas, que as diferenças entre o funcionalismo e a teoria da identidade são exageradas. Ele argumenta que sua *fórmula tópico neutra*, “O que está acontecendo em mim é como o que acontece em mim quando...”¹², é muito semelhante a uma descrição funcionalista, com a diferença de que esse tipo de descrição não trata de coisas tão abstratas como funções, mas de algo que está ocorrendo¹³, algo concreto. Essa é, inclusive, uma crítica que ele faz a Hilary Putnam, por este enfatizar o grau de abstração da linguagem mentalista, ao passo que para Smart ela parece ser inapropriada na descrição de sensações diretas (Smart, 1991).

De acordo com sua concepção de funcionalismo (Smart, 1995b), este faria uma descrição de um estado mental em termos de estímulos e respostas comportamentais (*inputs* e *outputs*). Esta definição de funcionalismo não é muito precisa, pois faltam as relações com e entre os estados internos, ou seja, os estados mentais, além das relações desses estados entre si. A essa estrutura proposta por Smart, da forma como aparece, é comumente atribuída ao behaviorismo. Mas, em outras passagens, ele acrescenta as relações entre processos e estados, formando o esquema funcionalista de *inputs*, relações entre estados mentais e *outputs*.

Aparentemente, na relação entre teoria da identidade e funcionalismo, Smart estabelece uma divisão de tarefas segundo a qual à teoria da identidade caberia identificar contingentemente esses estados mentais, descritos anteriormente em termos funcionais, a estados cerebrais. Essa divisão de tarefas é, na verdade, a estrutura do livro *A Materialist Theory of the Mind*. A segunda parte deste livro é considerada uma análise conceitual de conceitos mentais, não implicando na verdade da teoria da identidade, como pode-se ler em sua página 355: “Exceto por observações incidentais, nada que falei na Parte Dois *acarreta* a

¹² “What is going on in me is like what goes on in me when...”

¹³ Essa fórmula neutra será estudada mais adiante no capítulo 7.

verdade de uma teoria materialista da mente. O melhor que pode ser dito da Parte Dois é que nela o caminho foi suavizado para uma tal identificação.”

Também em Smart (1991 e 2006), ele expressa novamente a opinião de que algumas pessoas pensam que o funcionalismo suplantou a teoria da identidade e rebate dizendo que há pouca diferença ontológica entre as duas teorias, sustentando que “a maioria dos funcionalistas estaria feliz em dizer que as bases categóricas dos estados e processos funcionais são estados e processos cerebrais puramente físicos”.

Em outra passagem do artigo “Identity theory of the mind”, Smart aponta o fato de que assim “como Lewis e Armstrong, os funcionalistas definem estados mentais e processos em termos de suas relações causais com o comportamento, mas não vão adiante para identificá-los com processos e estados cerebrais, suas realizações neurais”. Desse modo, ele aponta o funcionalismo como uma etapa para a identificação entre mente e cérebro, de tal forma que essa teoria já está presente nos filósofos que postularam essa identidade. Posteriormente, no mesmo artigo, ele complementa que a identificação funcionalista por meio dos papéis causais leva à identidade dos mesmos com estados e processos neurais, dado que são eles que desempenham os papéis causais apontados pela teoria. Assim, no mesmo parágrafo, Smart afirma que, dependendo de como se define o termo funcionalismo, as teorias de Place, Armstrong e a sua própria teoria podem ser consideradas funcionalistas.

Esse processo de identificar os estados mentais após serem descritos funcionalmente parece ter acontecido essencialmente por meio da influência que Armstrong exerceu em sua filosofia, pois foi este filósofo quem levou Smart a identificar não apenas sensações e percepções com estados e processos cerebrais, como também crenças e desejos, ou seja, *todos* os estados e processos mentais. O processo se dá com a identificação dos conceitos de crenças e desejos – não está claro se também estão sendo utilizados aqui em um sentido próximo de atitudes proposicionais – com estados mentais, os quais são identificados contingentemente com estados cerebrais. Esse processo pode ser confirmado nas páginas 90-91 do livro *A Materialist Theory of the Mind*, onde Armstrong afirma que o argumento da teoria do estado central (*central-state theory*) tem dois movimentos distintos. O primeiro é caracterizado por uma análise lógica dos conceitos mentais, um estudo causal, que não acarreta nem exclui o materialismo; e o segundo trata de identificar esses estados mentais com estados físico-químicos do cérebro. Em seguida, afirma que se trata de uma identificação contingente ou científica.¹⁴

Enquanto Smart, como teórico da teoria da identidade, afirma que não é o caso de

14 Também para Place e Smart essa identidade é produto científico.

haver um tudo ou nada entre o funcionalismo e ele mesmo, Ned Block (1980) vai classificá-lo como um funcionalista de origem behaviorista, incluindo também nessa classificação Armstrong e David Lewis, que defenderam em algum momento uma forma de teoria da identidade. Armstrong (1993), por exemplo, defendeu o materialismo do estado central, teoria que identifica estados mentais com estados do sistema nervoso central, o que pode ser considerado um outro nome para teoria da identidade, sendo uma espécie de desenvolvimento desta teoria. Posteriormente, o próprio filósofo considerou a si mesmo e às ideias do livro *A Materialist Theory of the Mind* como funcionalistas causais. Apesar disso, por mais que faça sentido classificar Smart como funcionalista, ele próprio continuou denominando-se um “teórico da identidade” e defendendo essa teoria, mostrando suas semelhanças e diferenças com relação ao funcionalismo.

No livro *Nosso Lugar no Universo*, páginas 168-9, Smart apresenta as seguintes posições com relação ao funcionalismo, especificamente aquele desenvolvido por Putnam:

É claro que não conhecemos os detalhes dos nossos “estados de máquina” internos: esses estados são selecionados e separados pelo comportamento que tipicamente os origina e pelos estímulos ambientais que os provocam. Mas o teórico da identidade também seleciona estados e processos mentais exatamente da mesma maneira. (O funcionalista talvez também possa incluir nesta história as relações existentes entre os estados e os processos mentais, mas o teórico da identidade também pode fazer a mesma coisa, visto que costuma individuar os processos mentais pelas semelhanças e diferenças que estes apresentam entre si. Falando metaforicamente, acho que o funcionalista normalmente está no fundo de uma teoria “sinal-sinal”: a doutrina é apresentada como compatível com o fisicalismo e os funcionalistas têm o mesmo motivo para achar que um estado funcional está materializado fisicamente como uma identidade, e os teóricos da identidade também o têm. O funcionalista está interessado em dar ênfase às diferenças que existem entre os nossos *conceitos* do que é mental e do que é físico. O teórico da identidade é capaz de concordar com o seguinte: o nosso conceito comum de linguagem do que é mental não é um conceito fisiológico. E um conceito ‘neutro’: não é especificamente fisicalista. [...] Nesse sentido o teórico da identidade poderia concordar em dizer que o nosso *conceito* (comum) do que é mental não é fisicalista, é neutro, e ao mesmo tempo sustentar que o mental propriamente dito (processos e estados) é físico. (Smart, 1995, pp. 168 – 9)

Nesta citação, pode-se apontar as seguintes posições: (i) que a forma como o teórico da identidade relaciona os estados mentais a estímulos e respostas é a mesma como o funcionalista efetua, os estados mentais são selecionados em função dos estímulos e respostas que sofrem ou ocasionam no meio no qual está localizado; (ii) ambos também podem incluir as relações entre os estados e os processos mentais; (iii) a teoria sinal-sinal (do inglês *token-token*, que, nesta dissertação, será traduzido como instância/instância) compatível com o funcionalismo (alguns funcionalistas são denominados também fisicalistas, talvez a maior

parte deles) e apresentada pelo funcionalista, faz com que ele tenha os mesmos interesses que os teóricos da identidade em que esses estados sejam materializados fisicamente no mundo; (iv) os teóricos de ambas as teorias podem concordar quanto à distinção dos conceitos mentais e físicos.

Em seguida, ele continua afirmando no parágrafo seguinte (p.169):

O teórico da identidade também poderia concordar com o funcionalista quando este diz que há vários meios físicos pelos quais é possível perceber estados mentais definidos funcionalmente. Podemos esperar que neurologicamente as sensações de dor, digamos, de um visitante do espaço que veio de Alfa Centauro fossem diferentes das nossas. Pode ser assim até entre a dor que você sente e a que eu sinto. Mesmo assim o teórico da identidade poderia esperar algumas semelhanças relativamente abstratas de um tipo físico, exatamente como há no caso de um amplificador que usa transistores, de um lado, e um amplificador que usa válvulas, de outro. No entanto, o teórico da identidade sustentaria que existem identificações “tipo-tipo” ao menos se estiverem confinadas ao que sente uma pessoa em particular, e certamente isso é muito plausível. (Smart, 1995, p. 169)

Desse modo, Smart afirma (v) o acordo entre as duas teorias, quanto à múltipla realização de estados funcionais, ou pelo menos sua possibilidade, que é importante porque este aspecto é tradicionalmente utilizado contra a teoria da identidade; (vi) a teoria da identidade defende que haveria algum tipo de semelhança entre as realizações de um mesmo estado funcional; e (vii) defende graus de identidade “tipo-tipo”, por exemplo, relacionando os estados mentais de um mesmo indivíduo, ou relacionando os estados de indivíduos ou espécies diferentes. Smart conclui estes dois últimos aspectos da seguinte forma (p. 169):

Deveríamos esperar mais semelhanças entre a sensação de dor de Smith e a de Jones do que semelhanças entre os estados de crenças e desejo de Jones e de Smith. Deveríamos esperar ainda menos semelhança neurológica entre o estado cerebral de um gato correspondente a uma crença ou desejo e o mesmo estado de um cão ou de um cérebro humano. (Smart, 1995, p. 169)

É dessa forma então, que Smart lida, no livro *Nosso Lugar no Universo*, com as relações entre a teoria da identidade e o funcionalismo. É esperado que haja, pois, (a) mais semelhanças entre o mesmo tipo de estado mental em indivíduos com a mesma aparelhagem física, que (b) entre estados mentais diferentes nesses mesmos indivíduos. Além disso, haveria (ou poderia haver) (c) menos semelhanças que em (b) entre o mesmo estado mental realizado em aparelhagens físicas muito diferentes.

C. V. Borst (1970), na introdução ao livro *The Mind-Brain Identity Theory*, dirá que

materialismo de estado central é o nome para algumas versões da teoria da identidade. David Armstrong dirá, também, que os artigos “Is Consciousness a Brain Process?”, de U. T. Place, e “Sensations and Brain Process”, de J. J. C. Smart, “são exposições ‘clássicas’ do materialismo de estado central” (ARMSTRONG, 1993, p. 79). Isso mostra que a teoria foi sendo desenvolvida sem qualquer problema como uma forma de teoria de identidade, sendo considerada dessa forma tanto para os filósofos que teorizavam quanto para aqueles que comentavam essas teorias.

São justamente essas confusões entre as duas teorias que interessam neste trabalho. Porque, a princípio, elas são apresentadas como teorias distintas e conflitantes: por exemplo, o funcionalismo refutaria ou levantaria objeções à teoria da identidade. Assim, podemos dizer, por um lado, que o funcionalismo define estados, eventos e processos mentais em termos das relações causais entre um estímulo, relação com outros estados mentais (que também apresentam relações entre si) e uma resposta. E a teoria da identidade diria apenas que os estados e processos mentais são idênticos a estados e processos cerebrais¹⁵. Desse modo, a descrição de um estado mental em uma teoria diferiria da outra, pois uma seria dada em termos puramente relacionais e a outra em termos físicos. Apesar disso, em Smart, (1991, p.168), há a afirmação de que o teórico da identidade também pode tratar os estados e processos mentais por meio de suas relações, assim como faz o funcionalista.

Mas o que na verdade acontece com Smart, Armstrong e Lewis é que eles abdicam de uma descrição física de estados mentais em prol de uma que seja neutra com relação ao dualismo e ao materialismo. Ou seja, o realismo mental, presente em Smart, faz com que seja legítimo fazer relatos sobre as próprias experiências internas, porque são reais, mas com isso não se está falando sobre a natureza desses estados (Armstrong, 1993, p.121). O fato de a linguagem que é utilizada ser neutra é muito importante para a teoria da identidade e é ressaltada desde o artigo “Sensations and brain processes”. E é essa neutralidade que deve ser apontada como principal relação com o funcionalismo, pois possibilita a aplicação dos conceitos mentais definidos na teoria da identidade para ser identificada com outras entidades, quer sejam físicas, quer sejam imateriais.

Pode parecer que a identidade de estados e processos mentais com os estados e processos cerebrais é precedida por uma descrição funcionalista, mas vale salientar, para contrapor e complicar a análise em questão, que em Smart (2006, p.158) a experiência é considerada como algo existente apenas porque ela é um processo cerebral e, como este

¹⁵ Nos relatos de experiência, Smart também faz uso de relações com estímulos e/ou respostas no sentido funcionalista.

existe, assim deve ser também com aquela. Em outras palavras, o estado ou processo mental existem antes de sua descrição funcional. Portanto, embora haja essa proximidade com o funcionalismo, seria muito esperar que algo que fosse descrito apenas funcionalmente gozasse de existência para um teórico da identidade.

Também não ajuda muito fazer a distinção entre teoria da identidade tipo-tipo (cada espécie de estado mental é identificado com uma espécie de estado ou processo cerebral) e teoria instância-instância (onde cada estado mental individual/singular é identificado com um estado físico individual/singular), como vimos acima, afirmando que só o funcionalismo permite a identidade de estados mentais e físicos de tipo instância/instância, porque, no fim, Smart também defende uma identidade tipo-tipo. Chega a afirmar que é normalmente classificado como defensor da identidade de tipo, mas respondeu em Smart (1995) com um “sim e não” à pergunta de ser um teórico da identidade tipo-tipo, justificando que isso depende do grau de similaridade permitido. Em Smart (1991) ele afirma que as diferenças entre a teoria da identidade tipo-tipo e a instância/instância não são decisivas quando se trata das principais questões ontológicas, embora não afirme quais são elas.

Estas são passagens nas quais há menção de forma mais explícita à relação entre funcionalismo e teoria da identidade, mas pode-se, o que não é o objetivo deste capítulo, apontar quais posicionamentos da teoria da identidade a aproximam de uma teoria funcionalista. Pois há outras passagens que não foram relatadas aqui. Especialmente importante é o livro *Philosophy and Scientific Realism*, pois há muitas informações neste livro que poderiam caracterizar Smart como um funcionalista e, além disso, muitas das ideias apresentadas por Lewis e Armstrong, que são considerados funcionalistas até por eles mesmos, (pelo menos no caso de Armstrong), foram provenientes dessa obra. Por isso é importante fazer uma análise das ideias e relacioná-las com o que Smart escreveu antes e depois.

Para finalizar este capítulo, farei uma transcrição do parágrafo de *A Materialist of the Mind* no qual Armstrong fala de sua teoria, ilustrando o seu desenvolvimento, a relação com a teoria da identidade e também com o funcionalismo:

Meu livro foi parte de uma segunda onda de contribuições ao tópico [o materialismo representado por Place, Feigl e Smart] que ocorreu nos anos 1960. Talvez por causa do modo que meu argumento foi tomado das posições de Place e Smart, posições que eram muito próximas umas das outras, sou usualmente representado na literatura como sustentando uma teoria da identidade, que foi o nome dado à teoria de Place/Smart. Nunca estive muito entusiasmado sobre a frase e não a usei para

descrever meu ponto de vista. Como eu a via então, e ainda a vejo bastante, meu argumento tinha (e tem) dois passos. Primeiro há a proposta de uma análise lógica ou conceitual dos conceitos mentais. No Capítulo 6, §VIII, isto é chamado de *Análise causal* [*Causal analysis*]. A grosso modo, a asserção é que estados mentais são estados aptos para a produção de certas variedades de comportamentos e, em alguns casos, aptos para serem produzidos por certas variedades de estímulos. Depois eu apenas a chamei de teoria causal (veja meu artigo ‘The causal theory of the mind’, 1977). Atualmente, algumas vezes falo dela como a versão causal do funcionalismo. O segundo passo foi perguntar o que de fato é isso que é apto a causar e ser causado desse modo. A resposta para esta pergunta, eu respondi, é estados físicos do cérebro, assim alcançando o que eu chamei, seguindo Feigl, Materialismo de Estado-Central. (Armstrong, 1993, pp. xiii – xiv)

Surge então a questão de até que ponto estas teorias estão emaranhadas. Armstrong admite ter nomeado sua teoria de “materialismo de estado central” e no corpo do texto diz que tanto o artigo de Smart quanto o de Place são textos clássicos da teoria de estado central. Sua teoria também lança a hipótese da identidade entre mente e cérebro (identidade entre estados e processos mentais com estados e processos do sistema nervoso central, indo além ao incluir todos os estados mentais nessa identificação, também contingente. As semelhanças são muitas e eles estão muito relacionados entre si. Se o objetivo do trabalho for alcançado, haverá, no final, um mapa claro ou pelo menos um esboço mais fidedigno dessas relações. Talvez um ponto que distinga esses textos de Smart de Armstrong seja a ênfase na causalidade. A teoria de Armstrong está sendo apresentada, por ele mesmo, como uma versão causal do funcionalismo, porém em algumas passagens dos textos de Smart há opiniões suas contrárias à causalidade, ao ponto de vista causal na filosofia da mente.

3 FISCALISMO

Um esclarecimento que é importante para o entendimento da teoria da identidade, que é objeto deste estudo, diz respeito à determinação do que é físico, já que essa teoria é uma teoria fiscalista. O termo “fiscalismo” foi cunhado por Otto Neurath nos anos 1930 e também por Rudolf Carnap, ambos membros do Círculo de Viena. Ele se referia a uma tese linguística que afirma a tradutibilidade de qualquer afirmação em termos de uma linguagem de objetos físicos apenas. Na filosofia da mente contemporânea, o termo tem um significado diferente, sendo definido como a tese metafísica de que todas as coisas que existem são físicas. Tendo, portanto, uma abrangência ontológica.

Um aspecto que pode ser adiantado quanto à forma de fiscalismo adotada pelos filósofos da mente diz respeito à cientificidade, ou seja, à grande valorização da atividade científica. Isso pode ser visto em vários de livros e artigos da área, especialmente um cientificismo ligado à física, mas que também inclui a biologia, tendo a teoria da evolução presença confirmada nos artigos da maioria desses filósofos, além da neurofisiologia, cujo desenvolvimento alcançado no século XX foi o motor para o desenvolvimento da teoria da identidade e de outras teorias da mente durante todo esse século. Quanto a esta relação dos filósofos da identidade com a ciência, sobretudo com a teoria da evolução, John O'Connor (1969, pp. 3-4), falando de forma generalizada desses filósofos, enfatiza a consideração evolucionista que coloca os homens e os insetos dentro do mesmo gênero de comportamento. Além disso, também menciona a aproximação dos computadores da racionalidade humana. São, então, ramos diferentes das ciências e das técnicas que servem de inspiração para estes filósofos. O'Connor tem uma maneira simples de falar e, quando fala do fiscalismo, afirma que ele sustenta duas opiniões (1) de que os homens consistem dos mesmos materiais que a natureza, aqueles estudados particularmente pela física, e (2) que uma teoria científica pode dar conta completa e adequadamente da explicação dos seres humanos.¹⁶ (O'Connor, 1969, p. 5)

Ainda quanto ao cientificismo, Place inicia o artigo “Is Consciousness a Brain Process?” dizendo que a tese da identidade entre consciência e processos no cérebro, que ele está apresentando no artigo, é uma hipótese científica razoável. Também Smart e Armstrong

¹⁶ Apesar de fiscalismo e materialismo (em sentido amplo) serem sinônimos, o fiscalismo vai além da doutrina clássica de materialismo porque a Física que lhe confere conteúdo explica o mundo por meio de outros elementos além da matéria, como forças e campos, por exemplo. Os filósofos considerados nesta dissertação não ignoravam os elementos postulados por essa Física.

considerarão da mesma forma suas teorias de identidade. Então, como filósofos, eles estão colocando hipóteses científicas, empíricas. Esta é uma evidência mais que suficiente para sustentar o caráter cientificista da teoria da identidade dentro da filosofia analítica. Ela não só legitima o trabalho científico como também precisa dele para ser comprovada.

O primeiro a ser considerado será Herbert Feigl que, tendo sido membro do Círculo de Viena, fez parte dos iniciadores da teoria da identidade da forma como ela foi desenvolvida no início da segunda metade do século XX. O estudo do fisicalismo entendido e defendido por Feigl será fundamentado no estudo da sua concepção do conceito “físico”. Um ponto que poderá ser observado é a influência, mostrada no que se segue, do aspecto da intersubjetividade linguística presente no positivismo lógico do Círculo de Viena que influenciou todo o trabalho de Feigl, especialmente os envolvidos com a filosofia da mente e com o conceito de físico.

Herbert Feigl, no seu artigo “The 'mental' and the 'physical'”, determinará o físico relacionando conceitos e leis à explicação e à predição de processos inorgânicos. Os conceitos e as leis físicas seriam *suficientes*, em princípio, para fazer essas explicações e predições. No entanto, Feigl adiciona uma condição que amplia o escopo do que é físico. Essa condição diz respeito ao requerimento do emergentismo para a explicação dos fenômenos orgânicos da vida: se ele não for requerido, então os conceitos e as leis físicas tornam-se suficientes para a explicação dos fenômenos orgânicos e inorgânicos.

Feigl distingue dois conceitos de física, sendo o primeiro, mais restrito, sinônimo de científico, denominado físico₁. Este conceito é caracterizado pela forma como a especificação de seu significado é feita, ou melhor, dos elementos que estão envolvidos nessa especificação. No caso deste conceito, é essencial o envolvimento de conexões lógicas, que Feigl considera como podendo ser necessárias ou probabilísticas, com a linguagem observacional intersubjetiva. Também estão incluídos nesse conceito os próprios termos dessa linguagem. Segundo Feigl, são exemplos de físico₁ os conceitos teóricos em física, biologia, psicologia e ciências sociais. Dessa forma, esse conceito estaria envolvido com ciências de objetos bem diferentes, como orgânicos, inorgânicos, psicológicos e sociais.

O segundo conceito de física, físico₂, é definido por Feigl como aquele suficiente para fazer derivação dedutiva ou probabilística das afirmações sobre o domínio inorgânico da natureza. Neste caso, o físico diria respeito apenas ao que não tem vida. Assim, o conceito físico₁ abrangeria também o conceito físico₂.

Com isso, segundo Feigl, os conceitos mentais poderiam ser do tipo físico₁, enquanto podem ser introduzidos por uma linguagem observacional intersubjetiva de vida comum, que

incluiria a linguagem observacional da ciência. Os significados desses conceitos seriam dados por meio de postulados e regras de correspondência, como acontece com o conceito de campo magnético na Física.

Entre os termos que caem sob o conceito físico¹ há também duas distinções: a primeira se dá entre os termos observacionais e os teóricos; e a segunda distinção se dá entre os termos teóricos quanto aos níveis que possam assumir. Um dos exemplos científicos apresentados sobre essa divisão de níveis é o da termodinâmica clássica e os conceitos da mecânica estatística ou molecular.

A influência dessa visão científica, física, na filosofia de Feigl pode ser exemplificada pela seguinte passagem sobre a subjetividade, na subseção A da seção IV do artigo “The ‘mental’ and the ‘physical’”, na qual ele afirma a submissão de todas as coisas que existem à confirmação intersubjetiva, sendo esta sujeição à intersubjetividade um aspecto essencial do programa de trabalho básico e das hipóteses de trabalho da ciência. Apesar disso, essa subjetividade absoluta ou privacidade podem ser consideradas como possibilidades lógicas antes de serem rejeitadas na visão de mundo científica.

Em um texto posterior de Feigl, “Mind-body, *not* a pseudo-problem” (1960/1970), Feigl apresenta o materialismo como sendo uma hipótese, assim como o fez Place em 1956.

Além disso, a relação entre uma ciência física e uma concepção dualista da mente, ou seja, enquanto oposição do mental ao físico, aparece na subseção B da mesma seção, na qual ele afirma que se o conceito de “físico” não for entendido como um gênero, tipo, parte ou aspecto da realidade, mas como um método, uma linguagem ou um sistema conceitual, então não há lugar para tal concepção.

O fisicalismo de Place também diz respeito ao fisicalismo da Física. Sua teoria da identidade é colocada como uma hipótese segunda essa ciência e também a utilização de exemplos e suporte teórico são provenientes dessa ciência.

Smart defende um fisicalismo científico realista reducionista, ou seja, afirma que todas as coisas que existem são coisas físicas. Defende a existência ou realidade das entidades que são postuladas pela Física, como por exemplo elétrons, campos, e também a realidade das coisas macroscópicas, como mesas, árvores etc. e também que todas as outras ciências podem ser reduzidas à física. Além disso, também defende um determinismo com relação ao nível físico macroscópico. Os livros *Nosso lugar no universo* e *Philosophy and scientific realism* apresentam demonstrações dessa postura.

Talvez não seja muito ousado dizer que o principal motivo para Smart defender um realismo fisicalista reducionista seja a simplicidade que uma tal visão de mundo acarreta, quer

dizer, uma simplicidade de entidades, especialmente relacionadas ao problema mente-corpo. Se essa afirmação não puder ser aplicada a toda ontologia defendida por esse filósofo, poderá pelo menos ser aplicada a esse problema. Para apontar passagens que possam apoiar essa opinião, basta observar o último parágrafo do artigo “Sensations and brain processes”, no qual ele expressa a simplicidade e a parcimônia como critérios de escolha entre duas teorias igualmente consistentes com os fatos. Esta afirmação também está presente em outros lugares como no livro *Philosophy and scientific realism*, no tópico “Philosophy as more than the elimination of nonsense”, onde ele afirma a importância da plausibilidade, simplicidade, arbitrariedade e navalha de Ockham em argumentos filosóficos.

Além da referência à simplicidade também estar presente no artigo “Sensations and brain processes”, uma outra fonte de referência sobre isso se encontra no artigo “Methodology and Ontology” (1989). Ao falar de sua concepção de metafísica, Smart afirma que a metafísica busca, por meio de clarificação conceitual, chegar ao “ponto de vista mais provável do mundo” (1989, p. 50).

Em seguida ele afirma que “a metafísica é o fim mais conjectural e conceitualmente interessante da ciência total”. Assim, não há um limite bem definido entre a ciência e a filosofia, fazendo com que se assuma que a filosofia pode criar hipóteses científicas bem como a ciência lidar com discussões e assuntos de metafísica. A relação entre filosofia e ciência assume inclusive uma dependência da filosofia em função da ciência em alguns aspectos, como a confirmação da verdade da teoria da identidade mente cérebro. Por exemplo, ele contrapõe a biologia humana, a psicologia neurofisiológica, e teorias da percepção interna e externa à *mirada* interna do fenomenologia. O mesmo se passa para a explicação da visão.

É claro que uma pessoa poderia afirmar a complexidade da Física fundamentando-se no zoológico cada vez maior de partículas e entidades físicas procuradas e “encontradas” pelos físicos, sem mencionar os elementos teóricos estranhos como anti-matéria e tudo mais. Para Smart, essa multiplicidade de novas partículas descobertas em laboratório é irrelevante para entender a matéria ordinária. Os desenvolvimentos nessas áreas não trarão mudanças para as asserções fisicalistas.

Em Smart (1978, 1995), coloca-se o problema da suficiência do estágio teórico atual da Física para dar conta dos fenômenos biológicos e mentais. Durante a história da Física, ela passa por desenvolvimentos e rupturas, como a mecânica cartesiana que foi substituída pela mecânica newtoniana, por exemplo. Dessa forma, alguns filósofos sugerem que também a física terá que acrescentar novos desenvolvimentos para que os fenômenos biológicos e mentais possam ser explicados fisicamente, de maneira completa. Smart argumenta contra

essa visão, vinculando seu fisicalismo à física “atual”¹⁷, ou seja, argumentando que uma nova revolução na física não é necessária. Isso não significa que ele negue que possa haver uma nova revolução, mas nega que ela seja necessária para explicações fisicalistas na filosofia da mente.

O primeiro argumento em favor desse ponto de vista, da independência da explicação fisicalista da mente de uma nova revolução fisicalista, é uma analogia da mente com um computador: uma explicação mental a partir do nível de rede neuronal seria análoga a um computador a partir de sua rede de transistores. O segundo argumento trata da explicação da estrutura citológica desses neurônios para o entendimento da forma como eles funcionam: essa explicação requer apenas conhecimento de história natural e aplicação de princípios químicos e de teoria da eletricidade. O nível no qual as explicações seriam dadas estaria, então, à parte daqueles nos quais Smart espera que as revoluções aconteçam, como o nível sub-atômico ou unificação da mecânica quântica com a teoria geral da relatividade. Ou seja, ele acredita que para a explicação dos fenômenos mentais seriam necessários apenas conhecimentos de bioquímica e biofísica, e que os princípios dessas áreas já estão completos, ou seja, a física da matéria ordinária está completa em seus princípios fundamentais.

O fisicalismo de Smart envolve uma cadeia de reduções (Smart, 1963, 1995), como pode ser inferido do que foi exposto nos últimos parágrafos, da psicologia à biologia, da biologia à química e da química à física. Neste universo, o homem nada mais é que um mecanismo físico de uma complexidade extraordinária. Trata-se então, de uma redução ontológica, já que na ontologia de mundo do fisicalista só há lugar para coisas físicas. O principal problema envolvido, apontado por Smart nessas reduções, especialmente da química à física, está na complexidade computacional para se pôr em prática tal redução. Trata-se então de um problema tecnológico e não teórico. Uma formulação dessa redução aparece sob a seguinte forma: “qualquer modelo de um conjunto de sentenças verdadeiras da química é o modelo de um conjunto de sentenças verdadeiras da física” (Smart, 1995, p. 109).

Na redução da biologia à química, a teleologia que pode ser atribuída aos sistemas biológicos, é considerada uma derivação da seleção natural, que atuou ao longo de milhões de anos, sobre as mutações e demais processos puramente físico-químicos e aleatórios. O núcleo central da biologia seria, então, o bioquímico (e também biofísico), embora a biologia evolutiva também tenha um papel muito relevante. A relação entre essas disciplinas pode ser evidenciada na seguinte passagem:

17 Smart vinculou seu fisicalismo à Física do seu tempo e um fisicalista do século XXI vinculará seu fisicalismo à física de seu tempo.

A teoria da evolução está integrada na genética e a genética molecular está ligada à química. No entanto, de modo geral consideramos o princípio da seleção natural a base principal da evolução. (Smart, 1995, p. 119)

As dificuldades de se reduzir a biologia à química e a distinção entre biologia, química e física, foram mais cuidadosamente analisadas em Smart (1959a). Desde este texto, a biologia é comparada à eletrônica, por não ter leis gerais como a física e a química. Assim ele se expressa:

Assim como a engenharia de rádio usa a física para explicar por que um circuito com um certo diagrama de fios se comporta como se comporta, assim a biologia usa a física e a química para explicar porque organismos ou partes de organismos (e.g., núcleo celular), com uma certa descrição de história natural, se comporta do jeito que se comporta. [...] Biologia descritiva consiste em generalizações de história natural, não leis no sentido estrito. Então enquanto, falando grosseiramente, a engenharia de rádio é física mais diagramas de fios, então biologia é física e química mais história natural. (Smart, 1959a, p.364)

O texto do qual a citação foi retirada, “Can biology be an exact science?”, concorda com o capítulo III, “Physics and biology”, do livro *Philosophy and scientific realism*, questiona a possibilidade prática de se axiomatizar a biologia. Nesse texto, são apontados aspectos da biologia, distinguindo-se a sua metodologia da metodologia presente na física e na química. Além disso, é apontado o papel interteórico que a estatística desempenha na biologia, como na teoria da evolução e ecologia, que são ramos da biologia de caráter fortemente histórico. Claro que não só aí se tem a história natural, mas também no estudo das mitocôndrias e demais microestruturas celulares e mesmo nos estudos físico-químicos. A física e a química são usadas como ferramentas para explicar os *porquês* da biologia. Há, então, distinções metodológicas entre a física, a química e a biologia, mas Smart também apresenta uma opinião que persiste nos outros textos e que faz parte de sua tese fisicalista, que é a da inexistência de divisão de natureza entre os objetos dessas três áreas, ou seja, entre os objetos físico-químicos e os biológicos, sendo que estes são apenas muito mais complexos.

Para concluir, Smart considera que a psicologia é redutível à biologia e que, assim como esta, não possui leis, apenas generalizações, e que está sujeita à maioria das distinções metodológicas apontadas. Sendo que é incompatível com o materialismo, rejeita a emergência de leis ou propriedades irreduzíveis na biologia e na psicologia (Smart, 1970).

Armstrong e Lewis têm um ponto de vista semelhante com relação às ciências físicas, apesar de eles terem refinado mais a sua visão de mundo. Não nos deteremos aqui na

diferenciação de suas ontologias.

4 IDENTIDADES

Este capítulo aborda algumas teorias da identidade, principalmente a teoria contingente, visto que ela é a base sobre a qual se desenvolve a teoria da mente que é objeto deste estudo. Pretende-se partir da distinção entre tipos (*types*) e instâncias (*tokens*), abordar a identidade contingente, seguida pela crítica que Kripke desenvolveu para esta concepção, e abordar o que algumas vezes foi chamado de identidade ontológica, embora esta não seja uma teoria tão desenvolvida quanto a anterior. Fica em aberto o desenvolvimento de um tópico específico para a teoria necessária da identidade, se é que se pode chamá-la assim, ou então uma abordagem da lógica modal na qual essas formas de identidade sejam desenvolvidas.

4.1 Tipos e instâncias (*types* e *tokens*)

Há uma forma intuitiva de se distinguir *tipos* e *instâncias* na qual o entendimento é quase instantâneo. Considere a seguinte frase:

Rosa é uma rosa é uma rosa é uma rosa.¹⁸ (4.1.1)

quando se afirma que há nela três palavras, então está sendo utilizado um sentido de “palavra” como *tipo*, e essas palavras seriam: “rosa”, “é” e “uma”, que, neste caso, se repetem na frase. Mas em outro sentido, pode-se afirmar que há dez palavras, e aqui elas são tomadas como *instâncias*. Assim, intuitivamente, chama-se cada uma das palavras de tipo, e cada uma de suas *ocorrências*¹⁹ de instância, sendo o primeiro abstrato e *único*, e o segundo concreto e particular. Uma instância de palavra pode ocorrer como uma mancha de tinta em algum suporte, uma onda sonora, um sinal cinético ou visual, etc. Esta compreensão é intuitiva porque pode-se entendê-la sem ir mais a fundo na natureza de um tipo. Com isso, é possível apresentar outros exemplos para os quais a contagem de instâncias e tipos é feita corretamente. Isso mostra que a distinção entre tipos e instâncias, nesse estágio, faz sentido e

¹⁸ Tradução da frase “Rose is a rose is a rose is a rose.” do artigo “Types and tokens”, 2006, de Linda Wetzel e publicado na Stanford Philosophy Encyclopedia.

¹⁹ Segundo Linda Wetzel, não se deve confundir sinais com ocorrência, visto que nem toda ocorrência é um sinal.

é bem compreendida.

A partir do momento em que se consegue fazer essa distinção e se pergunta pela natureza do tipo, não há mais uma colaboração tão significativa da intuição como antes e não fica mais tão claro o que faz com que seja possível distinguir essas duas coisas, além do fato de tais objetos serem do mesmo *tipo*. Em outras palavras, se em uma mesa estão dispostas cinco maçãs e oito laranjas e é perguntado quantas frutas há na mesa, pode-se dizer que há treze, pois este é o número de objetos nesse ambiente que correspondem à palavra “frutas”. Mas pode-se dizer também duas, visto que há dois tipos de frutas sobre a mesa, maçãs e bananas. Geralmente se utiliza a palavra “tipo” quando se faz esse gênero de pergunta, “quantos tipos de fruta há na mesa?”. Isso é bem intuitivo e fácil de apreender, mas nada há nas coisas que vá além de seu nome e aspecto para permitir que tal procedimento seja levado a cabo. É porque se sabe que elas são do mesmo conjunto, partilham do mesmo nome ou possuem as mesmas ou grande número de propriedades, que se separa por tipos, e não porque todos sabem o que é o tipo ao qual cada uma delas corresponde, sem relação com os objetos concretos.

4.2 Estados, processos e eventos

Antes de iniciar o estudo da teoria da identidade mente cérebro, convém apresentar a definição de algumas palavras que são amplamente utilizadas tanto nesta teoria quanto no funcionalismo, quando se referem aos fenômenos mentais: estado, processo e evento. Para tanto, podemos recorrer às definições que David Armstrong oferece nas páginas 130 e 131 de *A Materialist Theory of the Mind*.

Chama-se de *estado* a situação que dura por um certo tempo, seja ele mais longo ou mais curto, e em cada instante dessa duração o estado existe inteiro. Armstrong dá como exemplo de estado físico o calor. Dizendo que se um corpo está quente por um certo tempo, ele estará quente em cada instante desse tempo. E de estado mental, a raiva. Estar enraivecido é estar inteiramente nesse estado em cada instante do período em que se está enraivecido.

Quanto ao conceito *processo*, Armstrong diz que um processo, ao contrário de um estado, não está completo em cada instante em que está ocorrendo, ele leva tempo para ser dado como completo, e a cada instante um montante se completa e outro resta para ser completado. Um processo é uma sequência de eventos. Exemplos de processos são uma

corrida, que é um processo físico, e um cálculo, que é um processo mental.

Já a palavra *evento* é usada por Armstrong nas situações em que um estado vem a ser ou quando ele passa (as mudanças de estado), ou o início ou o fim de um processo. Os exemplos para eventos dados são, ganhar uma corrida, evento físico caracterizado pelo fim de um processo, e tornar-se raivoso, a ocorrência de um estado mental.

John Heil (2004, 77-8), por sua vez, apresenta a definição de estado em termos de propriedades. Estado seria uma propriedade que uma substância possui. Mais precisamente (p. 99), ele define a palavra estado como uma propriedade que um objeto possui num determinado tempo. Eventos e processos seriam transições de estados. Um evento ocorre quando alguém passa a estar num certo estado. Processo seria uma sequência padronizada de eventos.

Estas não são as únicas abordagens e definições desses conceitos, mesmo no contexto de uma teoria da identidade entre mente e cérebro. Um bom exemplo pode ser encontrado em “Identity, Necessity, and Events”, escrito por Fred Feldman em 1974. Ele aborda três formas de teoria contingente da identidade psicofísica (ou seja, teorias da identidade mente e cérebro), que ocorreriam (i) entre substâncias, coisas ou “continuantes” físicos e mentais; (ii) entre eventos, acontecimentos e “ocorrentes” mentais e físicos particulares; (iii) e entre “tipos”, fenômenos e propriedades mentais e físicas. Ao fazer isso, Feldman apresenta três pontos de vista sobre eventos: (a) ponto de vista proposicional; (b) ponto de vista estrutural; (c) ponto de vista de eventos concretos. Neste artigo, “evento” aparece como um conceito mais geral do que nas formas especificadas por Armstrong, englobando o que este quer dizer por estados e talvez também por processos.

O primeiro ponto de vista (a) apresentaria três características: o evento é um objeto das atitudes proposicionais; não é necessária sua ocorrência para existir; e um dado evento pode ocorrer, deixar de ocorrer e voltar a ocorrer em outro momento. Este ponto de vista expressaria a identidade contingente segundo a forma: “um evento e é idêntico a um evento e' se e somente se é necessário que, para qualquer pessoa S e estado proposicional A , S tem A para e se e somente se S tem A para e' ” (Feldman, 1980, p. 149).

O ponto de vista estrutural afirmaria que eventos são entidades estruturadas complexas, formadas principalmente por propriedades, indivíduos e tempos. A identidade segundo esse ponto de vista ocorreria da seguinte forma: se “ a tem a propriedade de ser F em t ” é um evento, e “ b tem a propriedade de ser G em t' ” é um evento, então eles são o mesmo evento se e somente se os indivíduos a e b , as propriedades de ser F e de ser G , e os tempos t e t' , são idênticos.

O terceiro ponto de vista sobre eventos é denominado concreto. Ele pode ser formalizado da seguinte maneira: “um evento *e* é idêntico a um evento *e'* se e somente se alguma coisa é a causa de *e* se e somente se ela é a causa de *e'*; alguma coisa é um efeito de *e* se e somente se é um efeito de *e'*”. Feldman (1980, p. 151).

Embora possa-se fazer um estudo sobre as formas como estados, processos e eventos são definidos na literatura filosófica, especialmente na filosofia analítica, os conceitos que serão utilizados neste trabalho são os apresentados por David Armstrong, especialmente porque suas definições não foram problematizadas explicitamente pelos filósofos que interessam a este estudo.

Para concluir, segue-se uma citação com a forma como esses conceitos estão relacionados entre si na definição dos conceitos mentais feitas por Armstrong: “Observe finalmente que embora a causa mental que inicia e sustenta a deliberação é um *estado* mental, a ocorrência da própria deliberação não é um estado, mas um *processo*. Esse processo emerge em um *evento*: a tomada de uma decisão.” (Armstrong, 1993, 162)

4.3 Tipos de identidade

Um conceito essencial neste trabalho é o conceito de identidade. Afirmções de identidade são do tipo “*A é B*”, ou ainda “*A = B*”. O termo “*é*” na sentença “*A é B*” pode assumir algumas funções distintas. Com o conectivo “*é*” pode-se querer afirmar a existência de alguma coisa, como na sentença “Descartes *é*”; atribuir um predicado/propriedade a um objeto, como em “Platão *é* mortal”; subsumir uma classe em outra, como em “Homem *é* animal”; ou afirmar uma identidade como “2 *é* o menor número primo”.

O tipo de utilização do conectivo “*é*” de interesse neste trabalho é como função de identidade. Em teoria de conjuntos, dizer que o conjunto A é igual ao conjunto B é o mesmo que dizer que qualquer que seja o elemento, se ele é elemento do conjunto A, então ele é elemento do conjunto B e vice versa. Quando se considera que um objeto é um conjunto de propriedades, como muitos filósofos assim o consideram, então dizer que dois objetos são iguais é dizer que eles possuem as mesmas propriedades. Assim, suponha que Sócrates poderia ser o objeto que tem as seguintes propriedades {*é filósofo, bebeu cicuta, orientou Platão, é grego, é homem, é mortal*}, sendo estas as únicas propriedades de Sócrates. Podemos então dizer, Sócrates = {*é filósofo, bebeu cicuta, orientou Platão, é grego, é homem, é*

mortal}. Supondo que haja um objeto S do qual descobrimos que tem as seguintes propriedades: {é homem, é mortal, é grego, é filósofo, orientou Platão, bebeu cicuta} e nenhuma propriedade a mais, pode-se então dizer que $S = \{\text{é homem, é mortal, é grego, é filósofo, orientou Platão, bebeu cicuta}\} = \text{Sócrates}$, portanto, $S = \text{Sócrates}$. Mas se $S = \{\text{é homem, é mortal, é grego, é filósofo, orientou Platão, bebeu Coca-Cola}\}$, então a identidade não poderia ser afirmada.

Uma objeção com esse conteúdo será feita à teoria da identidade, ou seja, que as propriedades mentais não são as mesmas que as propriedades do cérebro e por isso essas duas coisas não podem ser identificadas.

Outra forma de identidade seria a translacional, que foi muito utilizada pelos filósofos do círculo de Viena. Ela é uma identidade ao nível da linguagem, ou seja, que evita a metafísica, embora isso seja questionável. Inicialmente, esses filósofos defendiam a identidade de sentido, ou seja, que qualquer sentença da linguagem poderia ser traduzida em uma sentença na qual só houvesse termos físicos, ou seja, que uma sentença poderia ser traduzida em outra sentença e o sentido seria preservado.

Além dessas, uma outra forma de se fazer a identidade é atentando para a modalidade, ou seja, se a proposição é necessária, possível ou impossível. Proposições são afirmações que podem ser verdadeiras ou falsas.²⁰ Nos restringiremos a lógicas bivalentes, cujas proposições só podem ser ou verdadeiras ou falsas, excluindo-se uma terceira opção. Proposições deste tipo podem ser representadas por meio de uma sentença de alguma linguagem, como uma das linguagens naturais, como francês, português ou inglês, por exemplo, ou então em uma linguagem formalizada. Assim, as sentenças “O gato é preto” e “*The cat is black*” são a mesma proposição: O gato é preto. Pode-se chamar essa proposição de q . Assim, quando se considera a proposição q , pode-se dizer que ela é verdadeira ou falsa. Podemos considerar uma segunda, chame-a de p , que afirma $2 + 2 = 4$, no sistema de números decimais. Também se diz que p é verdadeira ou p é falsa. Para atribuir um valor de verdade a q , em geral, olha-se para o mundo e, ao constatar que o gato é preto, diz-se que a proposição é verdadeira, caso o gato no mundo não seja preto, a proposição é falsa. Para atribuir um valor de verdade à segunda proposição, não é necessário olhar para o mundo²¹. Quando se atribui um valor de verdade às proposições p e q e se cria uma proposição composta $p \rightarrow q$, os valores de verdade

20 Em algumas lógicas elas podem assumir outros valores, como na lógica Fuzzy, onde a mesma proposição pode ser não só um pouco verdadeira, mas também bastante falsa, por exemplo. E uma hora você pode querer assumir uma proposição mesmo ela sendo só um pouco verdadeira, ou então, em outro momento, rejeitar uma proposição porque ela é muito falsa.

21 Para ser mais exata, olhar ou não olhar no mundo caracteriza as proposições como mundanas ou não mundanas.

encontrados são suficientes e necessários para a atribuição do valor de verdade para a nova proposição. Neste caso, se a proposição p é verdadeira e a proposição q é falsa, a proposição $p \rightarrow q$ será falsa, nos outros casos, será verdadeira.

O parágrafo anterior foi uma brevíssima introdução à lógica proposicional. Falaremos agora de uma extensão desta, que é a lógica modal. A modalidade diz respeito ao modo como as proposições podem ser verdadeiras ou falsas. Neste caso, segundo os modos, as proposições podem ser necessárias, possíveis ou impossíveis. Assim, poderia-se afirmar que a proposição *possivelmente* q é verdadeira quando q é verdadeira, mas e *necessariamente* q ? Tomemos a seguinte proposição: Gatos comem peixe. Esta proposição pode ser verdadeira para um determinado gato, para um número muito grande de gatos ou mesmo para quase todos os gatos, mas pode haver um gato que não coma peixe (pode ser apenas um único gato). A possibilidade de um gato não comer peixe faz com que a proposição: Gatos comem peixe, não seja necessária. Desse modo, a verdade das proposições modais não dependem da verdade de uma proposição particular, ou seja, uma sentença q pode ser verdadeira e mesmo assim a proposição *Necessariamente* q pode ser falsa. O mesmo vale para as proposições que são falsas. Neste caso, ele não será necessariamente verdadeira, mas poderá ser impossível ou possível. Um exemplo poderia ser: Tiririca é presidente do Brasil. Esta proposição é falsa atualmente, mas não há impedimento para que possa ser verdadeira no futuro. Ainda no sistema decimal, assume-se que *necessariamente* $2 + 2 = 4$ é verdadeira; e *necessariamente* $2 + 2 = 5$ é falsa ou $2 + 2 = 5$ é *impossível*.²² Nesta dissertação, $2 + 2 = 4$ será o exemplo de verdade necessária, pois este nível de modalidade se adequa ao utilizado pelos teóricos da identidade mente cérebro, especialmente J. J. C. Smart que, em seu artigo “Sensations and brain processes”, utiliza como a seguinte proposição para exemplificar uma proposição necessária: “7 é idêntico com o menor número primo maior que 5”.

A identidade também pode ser definida em termos modais, na forma: $A = B \rightarrow$ *necessariamente* $A = B$. Neste caso, não pode haver a *possibilidade* de A ser diferente de B . A possibilidade, como foi visto no parágrafo anterior, não depende estritamente da existência do objeto, porque mesmo que nenhum objeto tenha sido encontrado até então, ainda assim uma

²² Neste caso, não será apresentada a versão de lógica modal com mundos possíveis porque esse tipo de formulação e argumento não é utilizado nas críticas e formulações de interesse neste trabalho da teoria da identidade mente-cérebro. A crítica de Kripke pode ser encontrada no livro *Naming and Necessity* e em publicações do trecho desse livro no qual a crítica direta à teoria da identidade é feita presente em alguns livros de coleção de artigos.

Mundos possíveis podem ser vistos como formulações de *dicto*, embora chamar esse tipo de imaginação possa não ser a melhor palavra porque pode remeter a imagens visuais, uma definição mais abrangente de imagem que significasse ter ideias. Assim, mundos possíveis podem ser vistos como formulações em uma linguagem, seja ela lógica ou natural, de como o mundo poderia ser.

proposição pode ser *possível*. Se todos os corvos encontrados até hoje são pretos, nada impede que possa haver um corvo branco, ou seja, a possibilidade de existência de um corvo branco ainda pode ser afirmada.

Kripke (1972, p. 143), ao defender esse tipo de identidade, leva em conta que A e B são “designadores rígidos”. A definição de designadores rígidos leva em conta a utilização de mundos possíveis, que podem ser concedidos como existentes (Lewis) ou como modelos imaginários. Dizer que um designador é rígido é o mesmo que dizer que em todos os mundos possíveis, esse designador sempre designa o mesmo objeto, a mesma coisa. Assim, tomando o exemplo de verdade necessário dado acima, em todo mundo possível, $2 + 3 = 4$.

Todas as identidades envolvem designadores rígidos? Algumas formulações de identidade expressam que o termos da identificação não designam rigidamente e, portanto, a identidade estabelecida na proposição é *contingente*. Para conhecer o que designa uma proposição pode-se ou não recorrer à experiência. “Solteiro é não casado” é o exemplo de uma proposição para a qual não é necessário qualquer experiência, o próprio entendimento da linguagem na qual ela é enunciada já é suficiente para sua compreensão, porque a definição já está incluída no termo que está sendo definido. No entanto, para haver o conhecimento de outros tipo de proposições não é assim tão simples, pois envolvem experiência para comprovar a identidade. Este tipo de proposição é chamado a *posteriori* e dizemos que ela é *contingente*, porque o conceito que está sendo definido não contém a definição.²³ A proposição “A mente é o cérebro” é defendida como sendo um exemplo desse tipo de identidade. A proposição “Está chovendo agora” também expressa uma proposição contingente, pois é necessário fazer uma verificação empírica para saber se ela é verdadeira ou não.

Para se construir uma proposição que expresse uma identidade contingente é preciso que os termos de um lado da igualdade sejam diferentes dos termos do outro lado. Será mostrado nesta dissertação que, para construir a identidade mente = cérebro, os termos “mente” e “cérebro” devem ser definidos de forma independente, ou seja, nem a definição do termo “mente” deve conter o termo “cérebro” ou os termos pelos quais ele é definido, nem o termo “cérebro” deve conter o termo “mente” ou os termos pelos quais ele é definido. Além disso, o termo “mente” não deve estar contido no termo “cérebro” e, portanto, ser extraído por

23 Não há impedimento lógico para que uma proposição a posteriori seja necessariamente verdadeira, o que pode haver é uma limitação epistemológica. Com o desenvolvimento da nanotecnologia, por exemplo, uma alteração da estrutura atômica ou molecular do elemento Au fará com que o novo elemento ainda possua a mesma essência ou teria uma nova essência. Ou se a física mudasse todas as suas leis fundamentais e, com isso, as propriedades essenciais dos elementos da tabela periódica fossem alterados, o que dizer dos *novos* elementos?

meio de uma análise conceitual, nem o termo “mente” deve estar contido no termo “cérebro” e ser extraído do mesmo modo.

Tendo, então, definido os conceitos de interesse dessa forma, deve-se proceder a experimentos a fim de verificar empiricamente a identidade. Porém, a verificação de um termo científico empírico esbarra no problema da indução científica. A experiência em questão tem como objetivo verificar se a referência de um dos termos é a mesma que a referência do outro termo. Quando a identidade está em função da referência dos termos identificados ela é chamada de identidade *ontológica*, e geralmente tem inspiração na teoria desenvolvida por Frege.

Uma interpretação do tipo de identidade fregeana defendida por Smart e pelos outros teóricos que o seguiram é apresentada por Hintikka (2005, 128), que afirma:

Se o único propósito de identidades fossem marcar a identidade das respectivas referências de duas expressões, o status cognitivo das identidades poderia ser questionável. Eles poderiam, em última análise, meramente expressar que duas expressões se referem à mesma coisa. E a *propriedade* de tal identidade parece ser meramente uma matéria de linguagem, não da realidade à qual a nossa linguagem pertence.

Outro ponto de vista sobre a identidade contingente pode ser o de Feigl (1958), que utiliza um conceito de identidade no qual ela é uma relação. Esta se caracteriza pela designação de um mesmo indivíduo (ou universal) por diferentes rótulos ou descrições. Assim, um mesmo referente pode ser relacionado por dois rótulos diferentes, ou por um rótulo e uma descrição, ou ainda por duas descrições diferentes, por exemplo.

Embora Feigl aceite e considere que a identidade, quanto a seu significado, se dá segundo o princípio de identidade dos indiscerníveis, de Leibniz, ele também aponta dois tipos de identidades: lógica e empírica. A identidade empírica é subdividida em accidental, nomológica e teórica.

A última denominação de identidade a ser considerada não é muito precisa, chama-se identidade estrita. Ela foi incluída no artigo “Sensations and Brain Processes” para diferenciar o tipo de identidade que Smart estava defendendo, porém não foi definida e apresenta-se um pouco problemática. A dificuldade de entender o conceito de identidade usado por Smart é apontada por Norman Malcolm (1964/1969). Segundo ele, Smart teria dado exemplos que não evidenciam o caráter da identidade estrita que Smart diz adotar. O primeiro é que essa identidade não é uma continuidade temporal e espacial com algum fenômeno ou objeto,

também não é uma parte de um objeto, mas todo o objeto com suas partes. Este último aspecto pode ser apreendido como um objeto, do qual uma propriedade ou parte é salientada, é igual ao objeto do qual uma outra propriedade ou parte é salientada. Assim, se, para n sendo um número natural, um objeto O é composto pelas partes $(o_1 + o_2 + o_3 + \dots + o_n)$, então o objeto O cuja parte é o_1 é igual ao objeto O cuja parte é o_3 . Este é o sentido de identidade estrita que Smart apresenta.

No *Tratado sobre o entendimento humano*, Hume (1989) apresenta uma distinção entre fazer uma identificação vagamente e fazer uma identificação de forma estrita. Neste caso, então, a identidade estrita está em oposição à uma identidade vaga.

John O'Connor (1969) também dará uma definição de identidade estrita, que também dá a impressão de ser uma identidade forte, certamente algo mais forte que uma correlação. O'Connor afirma que “tradicionalmente [...] duas coisas são idênticas se e somente se elas partilham todas as suas propriedades não-modais. Algumas vezes isso é chamado de identidade estrita.”

5 TEORIA DA IDENTIDADE MENTE-CÉREBRO

A teoria da identidade mente-cérebro identifica contingente e ontologicamente estados, processos e eventos mentais a estados, processos e eventos cerebrais.

Essa teoria surgiu em oposição ao dualismo, que sustentava a existência de estados mentais puramente psíquicos, que não seriam redutíveis a estados físicos, e também como uma diferenciação em relação ao behaviorismo lógico, especificamente, o desenvolvido por Gilbert Ryle, por considerar a experiência mental como sendo algo real e identificá-la aos processos no cérebro. Inicialmente, essa diferenciação com relação ao behaviorismo ocorreu, especificamente, nos artigos de Place e Smart apenas no que diz respeito às sensações e às percepções, e posteriormente, com a Teoria de Estado Central de David Armstrong, foi estendida a todos os estados, processos e eventos mentais. Em razão destas relações, será feita uma breve exposição das teorias dualistas e behavioristas.

5.1 Dualismo

De forma resumida, a teoria dualista sustenta que a mente é distinta do corpo e que ela não pode ser explicada apenas em termos físicos. Tradicionalmente faz-se uma distinção entre o *dualismo de substância* e o *dualismo de propriedade* (teoria do atributo).

O dualismo de substância é caracterizado por considerar a mente como uma coisa não física que está conectada de algum modo com o corpo e que é responsável pelos estados e demais atividades mentais. Embora o conceito de substância possa ser bastante complexo, pode-se considerá-lo como o faz Armstrong, como aquilo que é logicamente capaz de existência independente. Este tipo de dualismo pode ser dividido em duas outras teorias, o *dualismo cartesiano* e o *dualismo popular*. O primeiro diz respeito à distinção que o filósofo francês René Descartes estabelece entre as coisas materiais, dotadas de extensão e posição no espaço, e a alma imaterial, que não possui nem extensão nem posição no espaço. O dualismo popular é a teoria do “fantasma na máquina”, onde o corpo humano seria uma máquina habitada por uma substância espiritual, não física mas com propriedades espaciais.

O dualismo de propriedade, por sua vez, é a teoria que considera que há apenas uma única substância, que possui tanto propriedades materiais associadas ao cérebro quanto

propriedades que são não-físicas, como sensações, pensamentos, desejos, etc. Um dualista desta teoria defende que estas propriedades não podem ser reduzidas ou explicadas em termos puramente físicos. São exemplos dessa teoria: o *epifenomenalismo*; o que Paul Churchland chama de *dualismo interacionista de propriedade*; e o dualismo de propriedade elementar.

Segundo o *epifenomenalismo*, os fenômenos mentais seriam produtos da atividade cerebral mas não exerceriam nenhuma atividade causal sobre os estados e processos cerebrais. Uma forma de explicitar esse caráter dos estados mentais considerados como epifenômenos, algo acima dos fenômenos, seria a seguinte citação de Paul Churchland:

Isso significa que a convicção universal de que nossas ações são determinadas por nossos desejos, decisões e volições é uma falsa convicção! Nossas ações são completamente determinadas por eventos físicos no cérebro, eventos que também causam os epifenômenos que chamamos de desejos, decisões e volições. (CHURCHLAND, 2004, p. 31)

Não fica difícil perceber que a teoria epifenomenalista leva a uma concepção de causação mental que é contraintuitiva. Embora se consiga fazer o caminho dos processos e estados cerebrais para os processos e estados mentais resultantes, o caminho inverso é negado e com isso toda uma tradição de psicologia popular perde seu referencial.

Ao contrário do epifenomenalista, o *dualista interacionista de propriedade* sustentará que, além de ser um produto da atividade cerebral, os fenômenos mentais também podem agir sobre a matéria. Assim, esta posição defende que há interação causal nos dois sentidos entre o cérebro e a mente.

Para a terceira forma dessa teoria, o *dualismo de propriedade elementar*, as propriedades mentais não teriam emergido, como ocorre nos dois casos anteriores, mas seriam irreduzíveis e não emergentes, como as propriedades eletromagnéticas em relação aos corpos com massa.

Alguns autores apresentam outras divisões para o dualismo, em alguns casos há apenas utilização de nomes diferentes para as mesmas teorias, outros fazem mais distinções entre os vários tipos, subdivisões, e com isso conseguem um número maior de teorias dualistas. Apresentaremos aqui mais algumas concepções classificadas como dualistas de propriedades. David Armstrong, selecionando dentre as teorias dualistas aquelas que consideram que os objetos físicos não têm natureza mental, ou seja, excluindo os mentalistas,

apresenta, além do dualismo cartesiano, o dualismo de feixe (*'bundle' dualism*), que diz respeito à descrição de Hume da mente como um feixe ou amontoado (*bundle*) de percepções, ou, como Armstrong define, uma sucessão de particulares não físicos ou itens distintos do corpo, embora lhe sejam relacionados.

Para Hume, as percepções são de dois tipos: as impressões, que se referem àquelas percepções mais vivas que ocorrem no momento em que as sensações ocorrem, assim, estando presentes “quando ouvimos, vemos, sentimos, amamos, odiamos, desejamos ou queremos” (HUME, 1989, [1710], p. 70); e as idéias ou pensamentos, que são definidas como percepções menos vivas, produtos da reflexão consciente sobre as impressões, quer com relação às impressões já vividas, a memória, quer com relação às projeções futuras ou imaginativas.

Além disso, Armstrong também apresenta uma classificação entre teorias dualistas *interacionistas* e *paralelistas* entre mente e corpo. Na primeira, o corpo mantém relações causais com a mente e esta com ele, como no dualismo interacionista de propriedades, descrito por Paul Churchland acima. Na segunda, o corpo age sobre a mente, mas esta não possui qualquer poder causal sobre o corpo. A versão paralelista da mente também pode afirmar que os estados mentais ocorrem junto com o estados cerebrais sem haver uma ligação causal entre eles, apenas acontece de ocorrerem junto.

Para não fazer apenas uma apresentação pontual das teorias dualistas da mente, convém considerá-las como o faz David Armstrong, apontando quais problemas estão envolvidos em algumas dessas teorias. Como esta dissertação não é uma defesa ou desenvolvimento da teoria da identidade, mas pretende mostrar como esta teoria e o funcionalismo se relacionaram ao serem desenvolvidas por determinados filósofos, não pretendo apontar os problemas de todas as linhas de dualismo, mas só aqueles que foram apontadas, neste caso, por Armstrong, que foi quem desenvolveu essa análise crítica de forma mais explícita e detalhada. Mesmo J.J.C. Smart, enfrentando um problema com o dualismo de propriedades já no artigo “Sensations and Brain Processes”²⁴, não fez uma análise dessas teorias.

24 Das objeções que Smart respondeu nesse artigo, a que teve as consequências mais relevantes para esta dissertação foi a objeção 3. Ela se inicia da seguinte forma: “Mesmo se as Objeções 1 e 2 não provam que sensações são algo mais além de processos cerebrais, elas provam que as qualidades de sensações são algo mais além de qualidades de processos cerebrais.” (Smart, 1959, p. 148)

5.1.1 Críticas ao dualismo de feixe

Armstrong fará uma crítica às teorias dualistas em três tópicos: o primeiro tópico faz uma crítica ao dualismo de feixe; o segundo é uma breve crítica ao dualismo cartesiano; por fim, no terceiro tópico, ele aponta dificuldades para qualquer teoria dualista.

Inicialmente, são apresentadas duas críticas ao dualismo de feixe. A primeira crítica é focada no princípio unificador da mente, para o qual Armstrong afirma haver quatro propostas de solução que ele analisa como inadequadas. O problema é apresentado da seguinte forma: como reunir todas as experiências que são minhas e, ao mesmo tempo, excluir todas as experiências que não são minhas. Que princípio permitiria fazer essa união? As soluções que ele afirma terem sido dadas como princípio unificador são: semelhança, causalidade, memória e continuidade.

A semelhança não seria uma boa solução porque há a possibilidade de existência de estados em mim que não se assemelham a nenhuma das minhas experiências anteriores, não tendo, portanto, como identificá-las como sendo da mesma mente. Além disso, há também a possibilidade de duas pessoas terem experiências muito semelhantes diante do mesmo objeto, por exemplo, sem que disso decorra que as duas experiências pertençam à mesma mente.

A causalidade relaciona experiências ao longo do tempo, ou seja, uma experiência atual é determinada por uma (ou mais de uma) experiência anterior. Esta causalidade seria interna, ou seja, ocorreria entre experiências mentais. Um contraexemplo para essa explicação seria uma experiência que fosse causada apenas por causas externas e, portanto, não estaria relacionada com a cadeia causal interna utilizada para unificar as experiências mentais. Além disso, ele também afirma o fato de as experiências de uma pessoa determinarem não apenas as suas próprias experiências como também as experiências de uma outra pessoa.

Com relação à memória como princípio de unificação da mente, ele reconhece o seu papel mas objeta por meio da ocorrência de um grande número (“a vasta maioria”) de experiências mentais que “passam como uma sombra”, que não envolvem memória de experiências passadas e sequer são recordadas posteriormente. Apesar disso, essas experiências são mentais. Dessa forma, Armstrong também rejeita a memória como princípio unificador da mente para o dualismo de feixe.

O próximo princípio unificador que ele considera para a teoria de feixe é o de *continuidade da experiência*, atribuído a John Mackie. A continuidade de experiência afirma que não há experiências discretas, mas um contínuo de experiência ao longo do tempo. Nessa

continuidade não há um limite entre os eventos mentais, eles se sucedem e se misturam. Esses eventos são discretos apenas por uma divisão arbitrária. Mais uma vez esse fundamento é rejeitado porque, para Armstrong, o que ele quer é justamente o que constitui essa continuidade da mente.

Além desses princípios, para os quais Armstrong apresenta objeções, ele também menciona um princípio mais forte, a unificação da mente por meio do corpo, que é apontada mas não abordada por ele. O motivo desse adiamento de abordagem é bastante claro, visto que está relacionado com as teorias materialistas e sua própria resposta ao problema corpo-mente. Um outro motivo é o rol de problemas com os quais essa proposta de unificação está relacionada.

Há também uma outra objeção à teoria de feixe que vai desempenhar um papel relevante na teoria de Armstrong e que subjaz às teorias dos outros materialistas australianos, ou seja, parece assumir uma importância implícita também nas teorias desenvolvidas por esses filósofos. Uma outra forma de unificar os estados mentais na teoria de feixe ocorre por meio da vinculação da mente ao corpo, através da afirmação de que a unificação dos estados mentais de uma pessoa é o corpo ao qual esses estados estão associados. Mas o princípio assim estabelecido também não é considerado satisfatório por Armstrong. A objeção a esse novo princípio de unificação diz respeito à *possibilidade lógica* de a mente subsistir em um *estado desencorpado*. Incluir o corpo na definição de mente, ou seja, como parte necessária dessa definição, será um impedimento para a possibilidade lógica da mente desencorpada. O que fundamenta esta objeção é relevante para o entendimento da relação entre o funcionalismo e a teoria da identidade. Abordaremos as implicações dessa possibilidade para a teoria da identidade e sua relação com o funcionalismo, isso será abordado em outro tópico. É suficiente destacar aqui que a distinção entre possibilidade lógica e possibilidade empírica: embora a mente desencorpada seja uma possibilidade lógica, Armstrong defende que ela é impossível empiricamente.

Além dessas objeções, Armstrong apresenta de forma generalizada uma explicação dos teóricos da teoria de feixe em termos de uma “relação única e indefinível” que ocorre apenas entre os itens de uma mesma consciência. Esta solução para o problema da unificação da consciência é interessante porque apresenta uma definição da mente em termos das relações desses estados, embora não sejam apontados os filósofos que fazem esse tipo de definição.

Finalmente, ainda com relação à teoria de feixe, há uma outra objeção que, como a anterior, é fundamentada no sentido da concepção ou definição de mente. Partindo-se da

suposição de que a mente é uma sucessão de itens não físicos, conclui-se pela possibilidade da existência independente, quer de um corpo material quer dos outros itens que fazem parte de uma coleção desses itens. Neste caso, ele interpreta Hume como um adepto dessa ideia, citando uma passagem do *Tratado do Entendimento Humano* no qual ele afirmaria que uma definição de substância à qual se atribui existência independente, “por si só”, faz com que todas as percepções sejam substâncias distintas. Veja que não é porque as percepções são substâncias que elas têm existência independente, mas porque as substâncias são concebidas como tendo existência independente que as percepções seriam substâncias. Assim, é assumido que as percepções existem por si sós, antes de qualquer consideração que as ultrapasse.

Observe que há uma semelhança entre as duas últimas objeções. Na primeira, Armstrong defende o sentido de mente desencorpada, ou seja, sua possibilidade lógica, e na segunda, nega que os itens mentais possam existir independentemente. A semelhança está em que em ambas as objeções está em jogo o sentido do que está sendo enunciado, embora no segundo caso seja considerada a existência e não apenas a possibilidade lógica. Quando se considera a mente desencorpada, não se excluem as relações que os seus itens mantêm entre si e mesmo com o meio externo. No segundo caso não se trata apenas de uma possibilidade lógica, mas da existência de itens mentais independentes. Armstrong, então, faz uma analogia desses itens com um sorriso sem face ou um soporífero sem algo que seja dessa forma. Assim como sorrisos não são capazes de existir independentemente das faces nas quais eles acontecem, um pensamento, uma dor, uma angústia não podem existir sem que exista uma outra coisa que as tenha.

5.1.2 Críticas ao dualismo de substância

A crítica ao dualismo cartesiano é breve e podemos considerá-la como crítica à solução dada ao último problema que vimos com a teoria de feixes. A crítica de Armstrong foi contra a existência de itens mentais independentes, que não seriam atributos ou propriedades de alguma coisa. No dualismo cartesiano, os itens mentais, pensamentos, são afecções da substância espiritual. Dessa forma, há uma unificação do “fluxo” de consciência, que não havia na teoria de feixe, e também um suporte para os itens mentais. Porém, o problema encontrado está na vagueza dessa substância. Não há, afirma Armstrong, outra coisa positiva que se possa dizer da substância espiritual além de que ela tem os itens mentais. Pode-se falar

dela negativamente, como dizendo que não tem propriedades espaciais, mas nada se pode falar de sua natureza positiva. Essa falta de especificação torna-se evidente quando se considera o problema do sono sem sonhos. Neste caso, como seria definida a substância? Que propriedades ela teria já que não tem mais as afecções mentais? O dualismo cartesiano não responde essa pergunta, o que o faz considerá-la como sendo vazia. A relevância dessa distinção entre as propriedades da substâncias e as propriedades mentais, de as duas coisas serem passíveis de ter suas propriedades especificadas separadamente, será vista especialmente quando a *teoria causal da mente* desenvolvida por Armstrong for considerada na seção 6.1.1.

5.1.3 Críticas a qualquer teoria dualista

As próximas objeções dizem respeito a qualquer teoria dualista e novamente é interessante notar que as objeções levantadas contra essas teorias não passam de formas subentendidas de formular a própria teoria de Armstrong, de preparar o terreno para a sua própria teoria como solução para os problemas da teoria da mente que são levantados e que não apresentam respostas satisfatórias por essas teorias. As *difficultades* são apresentadas na forma de questões, sendo elas: (a) “Um dualista pode dar conta da unidade entre mente e corpo?”, (b) “Como diferenciamos numericamente os objetos espirituais?”, (c) “A explicação dualista da origem da mente é plausível?”, (d) “Mente e corpo interagem?”.

Na dificuldade (a), o argumento apresentado por Armstrong se fundamenta na atribuição que ocorre no discurso cotidiano de propriedades mentais (espirituais) e físicas ao mesmo sujeito. Essa atribuição leva ao pensamento de que o homem é uma unidade e de que o corpo e a mente estão intimamente conectados. A teoria dualista então teria que dar conta dessa conexão, e responder à pergunta sobre a forma pela qual a mente e o corpo estão conectados. Para a solução desse problema, o argumento é construído com o postulado de uma “relação única e indefinível” que ligaria a mente ao corpo. Neste caso, as duas relações apresentadas que poderiam ser dadas como solução para este problema são: a *simultaneidade temporal* e o *relacionamento de causalidade*.

Essa proximidade entre corpo e mente não pode fazer uso da localização da mente dentro do corpo, como o argumento do piloto na nave. O próximo passo é uma sugestão de solução que poderia ser apresentada por um dualista que assume a seguinte estrutura:

Premissa 1: “É característico do nosso campo sensorial, em particular do nosso campo visual, que *nós* estamos no seu centro.”

Premissa 2: “Percebemos as coisas como mais próximas ou mais distantes de *nós mesmos*”.

Premissa 3: “Nosso corpo também é percebido como estando no centro de nosso campo sensorial.”

Conclusão: “Em algum sentido, parece que estamos localizados no mesmo lugar que nosso corpo.” (Armstrong, 1993, p. 26)

Esse argumento, que Armstrong sugere para o dualista, e que seria, caso o dualista siga os passos do argumento, uma forma de considerar o eu como sendo o próprio corpo como dado à percepção dos sentidos, leva à falsidade do dualismo, pois coloca o corpo e a mente em uma proximidade e mistura para a qual o dualista não tem uma explicação, visto que o dualista não pode simplesmente colocar a mente ou quaisquer de suas partes dentro do corpo.

A dificuldade (b) é aquela na qual os dualistas não conseguem encontrar um princípio de individuação entre substâncias espirituais ou coleções de itens espirituais que não seja de um modo negativo, ou seja, que seja um princípio de individuação não-espacial. Dado que o que diferencia dois objetos físicos é sua localização espacial ou espaço-temporal, como individuar duas substâncias espirituais se, por hipótese, elas não estão no espaço? Armstrong rejeita a distinção de história passada das substâncias espirituais, pois faz sentido a possibilidade de duas substâncias lógicas serem sincronizadas e terem o mesmo passado histórico espiritual ou tê-los idênticos por um certo limite de tempo.

O principal problema com as substâncias espirituais surge com a consideração da substância espiritual desencorpada, porque neste caso não há um corpo com o qual ela está relacionada e que possa lhe conferir individualidade. Ela deve ser uma possibilidade lógica para o dualista, dentro dos termos de sua própria teoria. Uma das questões colocadas é como distinguir entre *uma* substância espiritual que é relacionada igualmente a *dois* corpos, de um outro caso no qual *duas* substâncias espirituais idênticas são relacionadas a dois corpos. Armstrong apresenta este problema como não tendo uma resposta satisfatória por parte dos dualistas, por causa da forma como as coisas são individuadas no mundo, seus aspectos espaço-temporais, porque considera que qualquer que seja o objeto no mundo, como um fato, se ele é temporal também é espaço-temporal. Dessa forma, a não espacialidade da substância espiritual coloca em cheque a sua própria individualidade. É bom ressaltar que essa localização no espaço e no tempo é uma das características das coisas enquanto questão de

fato, não conflitando com sua possibilidade lógica.

A dificuldade (c) questiona se a explicação da origem da mente segundo os dualistas é plausível. Esta consiste de uma questão epistemológica que é abordada de acordo com o conhecimento científico disponível, ou seja, a explicação dada deve se adequar às teorias científicas vigentes. Neste tópico, Armstrong não usa referências bibliográficas, ou seja, não indica os nomes dos filósofos que está considerando em sua crítica, apenas utiliza uma noção generalizada de dualismo e analisa como a origem da mente seria definida nessa teoria, fazendo comparações com a teoria do funcionamento do cérebro dada pelo conhecimento científico. Desse modo, o ponto de vista dualista é posicionado especificamente com relação ao sistema nervoso tanto do homem quanto do animal, nos quais a mente teria surgido em um momento no qual há uma certa complexidade fisiológica é alcançada por esse sistema. Do ponto de vista atribuído ao dualista, a um certo nível de complexidade surge “uma entidade completamente nova, não-espacial”, que mantém uma relação completamente nova com o corpo, e esse surgimento não pode ser explicado por meio das “leis que lidam com as propriedades físicas das coisas físicas”.

O que está envolvido nesta dificuldade não é a plausibilidade lógica, visto que Armstrong defende que não há contradição lógica nessa teoria dualista, pois logicamente o sistema nervoso poderia dar origem a uma substância nesses termos dualistas. O que está em jogo é a plausibilidade diante dos conhecimentos científicos modernos. Como responder à questão “em que parte do desenvolvimento do organismo o corpo cria um objeto ou objetos espirituais?”, como responder essa pergunta mantendo a severa distinção dualista entre a matéria e a mente?

A próxima e última dificuldade para as teorias dualistas apresentadas nesta parte do livro por Armstrong é posta como a questão: “(d) Mente e corpo interagem?”. Esta dificuldade aborda o dualismo nas suas formas interacionista e paralelista e diz respeito a sua possibilidade empírica, pois se considera que não há problemas lógicos na interação entre mente e corpo.

Do ponto de vista interacionista, segundo o qual a mente e o corpo exercem mutualmente atividades um no outro, o dualista precisa satisfazer duas condições: a primeira é a presença de uma sequência iniciada com eventos no cérebro que antecedem ou causam os eventos mentais; a segunda tem a mesma estrutura, sendo os eventos físicos anteceditos ou causados por eventos mentais. Assim, o interacionismo discutido por Armstrong é bicondicional, as duas condições devem ser satisfeitas para que ele seja verdadeiro. Essa interação é determinada pelo vínculo causal entre um evento último de um tipo físico com um

tipo mental, ou viceversa, podendo haver uma cadeia de relações causais entre estados de um mesmo tipo antes de haver a interação entre os eventos mentais e físicos. Esse ponto de vista exige um ponto no qual ocorre essa interação. Descartes procurou o local de interação entre a mente e corpo no cérebro, localizando-o na glândula pineal. O problema epistemológico dessa concepção está em que não foi localizado um tal ponto de interação no cérebro. Além disso, Armstrong aponta que mesmo Sir John Eccles, neurofisiologista moderno que aceitou uma forma de dualismo interacionista, não desenvolveu uma teoria que desse conta de como essa interação ocorre, além de demandar um modelo de mente espaço-temporal.

Do ponto de vista paralelista, aquele no qual os eventos mentais estão correlacionados um para um com os eventos físicos, a interação entre mente e corpo se dá com o acompanhamento de um evento físico por um que seja mental, sem que este exerça qualquer relação causal com o anterior. A correlação entre esses eventos seria provida por um dicionário no qual cada evento mental poderia ser traduzido em seu evento físico correlato e vice-versa. O problema do paralelismo é a negação da ação da mente sobre o mundo físico. Um outro apontamento feito por Armstrong sobre a teoria paralelista é a consideração de que a interação entre corpo e mente é uma ilusão que surge da correlação regular entre eventos mentais e físicos, que assumiriam uma aparência de causalidade. Além disso, ele ressalta o princípio de ação e reação que domina o mundo físico, segundo o qual nenhuma matéria é passiva: quando um primeiro corpo atua sobre um segundo corpo, este também exerce uma reação, um ato, sobre o primeiro corpo. Portanto, o componente espiritual do homem, sendo passivo, seria diferente das outras coisas que existem no universo.

Dessas análises, Armstrong tira os requisitos para avaliação de uma teoria da mente. Assim, na página 36 do livro *A materialist theory of the mind*, ele lista os requisitos de uma teoria da mente. Desse modo, uma teoria da mente *deveria*:

- i. “permitir a possibilidade lógica da existência desencorpada da mente”;
- ii. “tratar acontecimentos mentais como coisas incapazes de existência independente”;
- iii. “dar conta da unidade entre mente e corpo”;
- iv. “prover um princípio de diferença numérica para mentes”;
- v. “ser cientificamente plausível”;
- vi. “permitir a interação causal entre mente e corpo”;

Uma outra condição que também é analisada na avaliação de uma teoria da mente é a

seguinte:

vii. “explicar o surgimento da mente no desenvolvimento do organismo”.

5.1.4 Críticas à teoria de atributos

Há também uma outra crítica à teoria de atributos, que Armstrong não considera junto com as teorias dualistas da mente mas que podem ser consideradas, como foi visto anteriormente, como uma teoria dualista de propriedades, pois considera o homem como uma única substância mas que possui propriedades físicas e propriedades mentais que não são redutíveis às propriedades materiais/físicas. A consideração dessas críticas, embora não estejam relacionadas com a teoria desenvolvida por Armstrong da mesma forma como estavam as direcionadas às teorias dualistas, é relevante porque a resposta à objeção 3 presente no artigo “Sensations and brain processes” de J. J. C. Smart, que era uma objeção de caráter da teoria de atributo, desencadeou o desenvolvimento de uma definição (ou relato) tópico neutra que esteve presente em todos os filósofos desse grupo desde então. Assim, para Armstrong, uma das maiores dificuldades da teoria de atributos é definir propriedades não-materiais e a forma como elas estão conectadas com o corpo.

Uma propriedade considerada irredutível por teóricos da teoria do atributo, como Brentano, é a *intencionalidade*, uma relação dos estados mentais com as coisas no mundo, quer existam uma árvore, um jantar, uma brincadeira de infância, quer não existam, como um unicórnio, um alienígena, uma experiência recordada que não existiu. Uma definição dada por John R. Searle (2002, p.1) para esse caráter de direcionamento é dada nos seguintes termos: “Intencionalidade é aquela propriedade de muitos estados e eventos mentais pela qual estes são dirigidos para, ou são acerca de, objetos e estados de coisas no mundo.” Apesar dessa definição, esse direcionamento ocorre mesmo quando não existem os objetos para os quais aponta no mundo. Além disso, a intencionalidade também pode aparecer como representação, além de ter outras peculiaridades como sua relação com os atos de fala, mas esses aspectos não são necessários no presente trabalho. É suficiente considerar seu aspecto de direcionamento.

Com a consideração da intencionalidade, as condições a serem satisfeitas por uma teoria da mente tornam-se acrescidas da condição:

viii. “explicar a *intencionalidade* dos estados mentais”.

A forma da teoria do atributo que Armstrong analisa mantém a identidade de mente e cérebro, ou seja, “como uma questão de fato empírica, a mente é o cérebro” (Armstrong, 1993, p. 46). As propriedades mentais, mesmo sendo irredutíveis à matéria, também são propriedades do cérebro. Dessa forma, considerando os estados mentais como atributos do corpo, a teoria do atributo consegue prover três das condições levantadas anteriormente: a *ii*, a *iii* e a *iv*. Apesar disso, outras condições apresentam dificuldades e respostas insatisfatórias. A primeira é a condição da possibilidade lógica de existência desencorpada. O movimento que Armstrong faz para mostrar que essa condição não é satisfeita é o seguinte: primeiro ele utilizou a condição *i*, segundo a qual estados mentais são incapazes de existência independente; em seguida fez a identificação das propriedades não-físicas postuladas pela teoria do atributo com esses estados mentais; chegando, assim, à conclusão de que uma coisa não pode ter apenas propriedades não-físicas.

Uma outra dificuldade é encontrada com respeito às condições *v* e *vi*, que são tratadas em conjunto porque a solução da *vi* implica na forma como a *v* se dá. Começando pela condição *vi*, exigência de uma teoria da mente permitir a interação causal entre mente e corpo, o argumento principal adotado por Armstrong toma a seguinte forma: (a) “se pensamos os estados mentais como um conjunto de atributos do cérebro, é muito difícil negar que o estado da mente é unicamente determinado pelo estado corrente do cérebro.” (Armstrong, 1993. p. 46) A possibilidade de que alguns estados da mente variem independentemente do estado físico corrente do cérebro leva a uma teoria paralelista da interação mente e cérebro para a teoria de atributo, o que exigiria um montante maior de argumentos para ser sustentada, por isso ela é rechaçada. (b) Nesta fase do argumento, o teórico da teoria de atributo, impelido pelos fatos empíricos, teria que afirmar que “a mente é o cérebro”, que “o mesmo objeto que tem as propriedades físicas de um cérebro também tem as propriedades mentais”, podendo ser descrito como “o cérebro” ou “a mente”. É interessante ressaltar que Armstrong aponta a semelhança entre a teoria de estado central, a sua teoria da mente, e a teoria de atributo, sendo que a distinção entre elas é dada pelo fato de a teoria de estado central só aceitar a existência de propriedades físicas, ou seja, rejeitar as propriedades não-físicas que são postuladas pela teoria de atributo. A conclusão desse argumento, a primeira, é a seguinte: (c) a teoria de atributo “permite a interação entre mente e corpo, enquanto ainda permite que as propriedades mentais sejam unicamente determinadas pelo estado físico corrente do cérebro.” (idem)

A segunda parte do argumento apresenta mais duas premissas/suposições: (d) “o

cérebro opera apenas de acordo com aquelas leis físicas, químicas e biológicas que explicam o trabalho do resto do corpo” (Armstrong, 1993, 47); (e) mostrou-se que “os trabalhos e poderes especiais do cérebro são simplesmente um resultado da tremenda complexidade de sua estrutura física, e de forma alguma um resultado de modos de operar radicalmente novos.” A segunda conclusão, forçada, seria que (f) as propriedades postuladas pela teoria de atributo seriam causalmente ociosas; ou que envolveriam a emergência de novas leis físicas.

Com relação ao problema epistemológico, a teoria de atributos não está de acordo com as descobertas ou as intuições da maioria dos pesquisadores modernos de neurofisiologia, que tendem a considerar o crescente número de evidências em favor de um funcionamento do cérebro em termos de leis físicas, químicas e biológicas, como acontece com o restante do organismo. Como alguns dos argumentos entraram como suposições e as evidências estão sendo construídas, não há uma resposta definitiva para a questão de se a teoria que ela é verdadeira ou falsa, mas se recorre a uma tendência da comunidade científica para apoiar a insuficiência da teoria de atributos.

Assim como para o dualismo, a teoria de atributo teria que explicar o surgimento dessas propriedades mentais não-físicas no desenvolvimento do organismo. Em que momento do contínuo desenvolvimento do organismo surge a intencionalidade? Neste ponto a teoria também parece não ter uma resposta satisfatória.

Um último argumento, e este se aplica a todas as teorias da mente que sejam não materialistas, é o argumento da supremacia da física. Trata-se de uma posição reducionista na qual a biologia é reduzida à química que, por sua vez, é reduzida à física. Não é difícil perceber como ela exerce sua ação nas teorias consideradas neste tópico, portanto visto que o fisicalismo será abordado em outro tópico, não será necessário explicitar o desenvolvimento da argumentação de Armstrong neste momento, mesmo que ele tenha considerado algumas objeções a essa redução e respostas a ela. Com isso, o tópico que trata das teorias dualistas ou não-materialistas da mente está encerrado e apontou as condições sob as quais a teoria de estado central foi estabelecida, como será mostrada mais à frente, na seção 6.1.

5.2 Behaviorismo

O behaviorismo é uma teoria materialista da mente que defende a explicação de todos os processos mentais em termos do comportamento físico e de tendências (disposições) para se comportar, apresentados pelos corpos. Harris Savin (1980) apresenta duas principais correntes dessa teoria, representadas por Carl G. Hempel e B.F. Skinner, que seriam o behaviorismo lógico e o behaviorismo psicológico, respectivamente. Hempel faz uso de uma tese epistemológica que determina a inteligibilidade de sentenças psicológicas através da descrição de comportamento físico: em outras palavras, quer dizer, as sentenças psicológicas são inteligíveis, porém descrevem comportamentos físicos. Já Skinner trata mais do procedimento de se fazer psicologia, afirmando que o único objetivo interessante da psicologia é prever o comportamento e para tanto é desnecessário que se considerem os estados e processos mentais; porém, ele não afirma que tais entidades mentais não existam. Mesmo com essa diferença de perspectiva, Savin aponta o fato de que tanto Skinner quanto Hempel partilham a convicção de que não há diferenças importantes entre pessoas e seres inanimados. Hempel (1935-1989, p. 15), inclusive, afirma explicitamente essa indistinção entre as ciências, apresentada pelo Círculo de Viena e da qual ele partilha, que se opõe à tese epistemológica que afirma haver uma “diferença fundamental entre psicologia experimental, uma ciência natural, e psicologia introspectiva; e em geral, entre as ciências naturais de um lado, e as ciências da mente e cultural de outro.” Ainda sobre o Círculo de Viena e sua influência na filosofia, especialmente a filosofia da mente, pode-se citar uma passagem de Hilary Putnam (1980. p. 25), na qual ele afirma:

Os positivistas de Viena, em sua fase ‘fiscalista’ (cerca de 1930) – tomaram o conselho de Russell tão seriamente a ponto de produzir uma doutrina que chamamos *behaviorismo lógico* – a doutrina que, assim como os números são (alegadamente) construções lógicas a partir de *conjuntos*, assim *eventos mentais* são construções lógicas a partir de *eventos comportamentais* atuais e possíveis.

Essa passagem de Putnam contextualiza o behaviorismo nas teorias que estavam se desenvolvendo no ambiente filosófico desse período. Ela ajuda a elucidar a relação que a filosofia analítica mantém desde então com a lógica na consideração dos problemas mente-corpo, especificamente. A influência do Círculo de Viena, da qual Feigl é apontando como um dos integrantes, se estende além da doutrina behaviorista e influenciará o desenvolvimento

tanto da teoria da identidade quanto até mesmo do funcionalismo, como teremos a oportunidade de acompanhar.

Além dessas duas correntes principais, há também mais uma corrente importante dentro do behaviorismo, o behaviorismo metodológico, que defendia uma psicologia que tivesse como objeto o comportamento e não a mente. Um psicólogo que exemplifica essa forma de behaviorismo é John Watson. Watson escreveu o livro *Psychology as the Behaviourist Views It* em 1913, livro que também é conhecido como o *Manifesto Behaviorista*. Outros exemplos das correntes de behaviorismo psicológico são Ivan Pavlov, Edward Thorndike.

A teoria behaviorista com a qual os filósofos considerados nesta dissertação dialogam com mais frequência, com exceção de Feigl, e da qual faziam parte é aquela desenvolvida por Gilbert Ryle, especialmente no livro *The Concept of Mind*. O behaviorismo defendido por Ryle está na mesma linha do de Hempel, ou seja, trata-se do behaviorismo lógico, e ambos defendiam um ponto de vista segundo o qual o significado de uma proposição é dado por suas condições de verificação, que deveriam ser condições físicas. Portanto, toda sentença psicológica teria o mesmo significado que uma sentença física (SAVIN, 1980, p.11). Como Smart expressa no artigo “Sensations and Brain Processes”, o relato de uma pós-imagem é uma tentativa de reportar uma experiência visual, ou então, “a expressão verbal de dor, substitui o choro, não o descreve.”²⁵ Ou seja, dizer que se está com dor é considerada como uma manifestação da dor, assim como o choro também é uma manifestação desse estado. Dizer que se está com dor e chorar são manifestações de dor. A preocupação principal de Ryle nesse livro era como correlacionar os conceitos mentais uns com os outros e com outros gêneros de conceitos.

Nas primeiras formulações do behaviorismo lógico, e aqui será utilizada principalmente a versão de Hempel, as afirmações sentenciais contendo termos psicológicos eram traduzidas ou retraduzidas em outra afirmação sentencial maior, na qual os termos mentais eram trocados por termos físicos, sem que houvesse perda de sentido. Dessa forma, os termos mentais seriam abreviações de termos ou descrições físicos. Mesmo na ciência, alguns conceitos são abreviações de uma descrição, mais extensa, contendo o mesmo significado que o conceito original. Neste caso, os conceitos podem ter diferentes descrições, ou seja, têm o mesmo sentido que diferentes sentenças físicas, sendo que o que vai determinar seu significado são as condições de verificação. Se existe uma condição para verificar uma sentença ou conceito, então essa sentença ou conceito tem conteúdo e é significativa, caso

25 Citação atribuída a Wittgenstein. § 244 das *Investigações Filosóficas*.

contrário, falta-lhe tanto o conteúdo quanto o significado.

O fundamento condicional da verificação das afirmações psicológicas legitima a *metodologia* adotada pelos behavioristas lógicos, ou seja, listar em que condições as afirmações psicológicas são verificadas. Assim, este se estabelece como o principal problema do behaviorismo. Um exemplo apresentado por Hempel e que pode ser de interesse neste ponto é a listagem de condições que podem verificar a sentença “Paul tem uma dor de dente”:

- a. Paul chora e faz gestos de tais e tais tipos.
 - b. À questão “Qual é o problema?”, Paul declara as palavras “estou com dor de dente.”
 - c. Um exame mais detalhado revela um dente cariado com raiz exposta.
 - d. A pressão sanguínea de Paul, seus processos digestivos, a velocidade de suas reações, mostram tais e tais mudanças.
 - e. Tais e tais processos ocorrem no sistema nervoso central de Paul.
- (Hempel, 1980, p. 17)

É justamente por esse tipo de tradução que os behavioristas são conhecidos entre os filósofos, quer para se aderir a essa teoria, quer para criticá-la. Observe também que as cinco condições elencadas são públicas, observáveis, e principalmente físicas. São testáveis. Esse é um critério científico do behaviorismo.

Em 1977, Hempel faz uma nota com caráter de prefácio para uma das versões de seu artigo, publicado inicialmente em francês em 1935, na qual afirma ter mudado a sua forma de fisicalismo da variante translacional para uma forma, se se pode dizer, de consideração reducionista, por passar a defender não a definição completa de propriedades e estados mentais, mas um ponto de vista no qual esses são parcialmente caracterizados pelas disposições comportamentais. Desse modo, essas descrições físicas podem ter mais ou menos o mesmo significado das sentenças psicológicas. Ou seja, ele não pretende mais definir os conceitos mentais apenas em termos de comportamentos, mas considerar estes como sendo apenas uma parcela do que constitui um estado ou uma propriedade mental. Não tenho informações sobre o que fez Hempel mudar de ideia, mas as formas como a teoria da identidade e o funcionalismo foram formuladas pelos filósofos considerados nesta dissertação está em conformidade com essa ideia que Hempel menciona, porque ambas as teorias utilizam os comportamentos como parte de sua definição de mente, mas não como o elemento único nessa definição.

Um ponto que pode ser controverso na relação entre essas teorias e o behaviorismo diz respeito ao fato de que alguns filósofos são considerados behavioristas, ao passo que esse

rótulo é rejeitado pelos próprios filósofos ou por seus discípulos, viz. Wittgenstein e Ryle. O último dirá, na página 327 do livro *The Concept of Mind*: “A linha geral deste livro será inofensivamente estigmatizada como behaviorista.” Aqui podem ser pensadas as distinções entre os diferentes tipos de behaviorismo. Com essa consideração, Ryle estaria em uma fase mais adiantada de behaviorismo; basta, para isso, observar os objetos de sua análise conceitual que entram nos títulos dos capítulos do livro: desejo, emoção, disposições e ocorrências, autoconhecimento, sensação, imaginação, intelecto.

Gilbert Ryle (1949, p. 319 e seguintes) inicialmente faz uma distinção entre a forma como a psicologia lida com seu objeto, que é o mesmo objeto para o detetive, o professor, o biógrafo, o economista, o historiador, o jogador, o empregado, o confessor, os pais, os amantes, os romancistas etc. Aponta que a principal distinção está em que a psicologia pesquisa as explicações causais para esse objeto. A psicologia está interessada na mente de uma forma que as outras pessoas e ciências, mesmo que também estejam interessadas pelo mesmo assunto, não estão. Essa é a visão geral para a qual é apontada uma exceção com os psicólogos que apenas fazem mensurações e coleções dessas mensurações. Estes psicólogos não estariam imediatamente preocupados com a causalidade, mesmo assim há uma esperança de que esses dados coletados servirão para construir correlações funcionais ou leis causais, sendo seus trabalhos uma fase preliminar e preparatória (Ryle, 1949, p. 327).

Ainda com relação à psicologia, há duas opiniões às quais Ryle se opõe. A primeira afirma que o objeto da psicologia é o “estudo empírico dos poderes mentais, das propensões e performance das pessoas”. A segunda, que só a psicologia pode dar uma descrição da mente apropriada por meio de seus termos técnicos (Ryle, 1949, p. 327).

O que Place e os demais filósofos materialistas farão é desmetaforizar a expressão “em minha cabeça”, a qual Ryle considera como uma metáfora²⁶, e cuja variante, “em minha mente”, seria dispensada por ele por dar a idéia de que a mente é um lugar estranho no qual seus ocupantes teriam estatuto de fantasmas. Essa desmetaforização não irá contra a concepção ryleana, porque este também rejeita a interpretação de “em minha cabeça”, no sentido de se encontrar uma melodia na cabeça de alguém por meio de uma cirurgia, mas enquanto as experiências mentais são estados e processos cerebrais. Segundo uma objeção comumente levantada contra a teoria da identidade, dizer que os estados mentais de um objeto redondo e da cor de uma laranja ou do azul do céu deveria ser encontrado como tal no cérebro.²⁷

26 Ver o subcapítulo com esse título no livro *The Concept of Mind*, Capítulo II, item (5).

27 É interessante observar que os fenomenalistas defensores dessa objeção não exigem que haja tal coisa no espírito, no máximo dizem que há formas dessas coisas no mundo platônico, mas mesmo assim, não exigem

Um outro conceito que Ryle desenvolveu e que tem bastante relevância na teoria da identidade e no funcionalismo causal é “disposição”. Este termo está relacionado com a tendência para um determinado comportamento e será desenvolvido melhor na seção 6.1.1.1. Nessa seção, tratando do funcionalismo causal de David Armstrong, será dada uma breve explicação sobre esse termo e sua diferenciação na teoria de estado central. A abordagem do behaviorismo presente no livro *A Materialist Theory of The Mind* é um pouco particular, pois trata da teoria de um ponto de vista de um teórico que tenta se apropriar de alguns conceitos e estabelecer novas relações e interpretações. Além disso, a estrutura e objeto do livro desse livro e do livro *The Concept of Mind*, de Ryle, é similar, pois ambos tratam de fazer um estudo conceitual dos conceitos mentais e os resultados alcançados por ambos podem ser comparados, para quem assim se dispuser. Armstrong examina dez pontos do behaviorismo. Alguns desses pontos serão considerados em algumas passagens da dissertação. A seguir será feita uma breve enumeração dos dez assuntos apontados: (1) o papel das disposições; (2) assimetria no conhecimento das disposições; (3) relação entre predicados e características comuns dos estados mentais que caem sob o mesmo predicado – noção de família; (4) casos centrais e casos periféricos de exemplos de um conceito; (5) consideração da quase impossibilidade de existir um comportamento completamente inambíguo para um conceito; (6) inexistência de uma divisa rígida entre as instâncias que caem e que não caem sob um conceito mental, especificamente; (7) pressuposições no uso de conceitos mentais; (8) contextualização temporal de um comportamento ao qual se aplica um conceito mental; (9) ausência de uma função descritiva ou puramente descritiva para um grande número de conceitos mentais; (10) impossibilidade de traduções de alguns conceitos mentais em termos físicos.

Fazendo novamente uma consideração sobre as condições que uma teoria da mente deveria prover, exigidas por Armstrong, a teoria behaviorista não apresenta problema em três dessas condições, *iii*, *iv* e *v*, que são, respectivamente, unidade mente-cérebro, diferenciar as mentes numericamente; e ter plausibilidade científica.

O behaviorismo apresenta algumas dificuldades relacionadas a: (i) possibilidade lógica da existência da mente desencorpada, já que não pode explicar a mente sem recorrer a definições e explicações da mente em termos de comportamento corporal; (vi) interação causal entre mente e corpo, pois aqui alega-se que os estados mentais não são considerados como partes causais da explicação da mente, e na caracterização do behaviorismo feita por Armstrong há ênfase na explicação por meio da disposição para se comportar, de forma que a

que seja confirmado.

mente não exerce um papel na cadeia de causalidade; (viii) insatisfação na explicação da intencionalidade; negação da existência de processos mentais internos. Não é difícil perceber e necessário explicitar que o behaviorismo não satisfaz o restante das outras dificuldades, *ii* e *vii*, que tratam, respectivamente, dos estados mentais como coisas incapazes de existência independente, ou seja, não são substância, e da explicação de como a mente surgiu no desenvolvimento do organismo.

O que foi apresentado sobre o behaviorismo é suficiente para um entendimento do contexto teórico no qual se desenvolveram tanto a teoria da identidade como o funcionalismo. A maioria dos filósofos considerados neste trabalho foram behavioristas e em certo momento no decorrer do caminho continuaram sendo behavioristas na consideração da mente em nível superior, como crenças, desejos etc. Além disso, como será mostrado, a estrutura funcional de explicação de alguns ramos dessas teorias é bem semelhante.

5.3 Teoria da identidade

A teoria da identidade entre mente e cérebro, que é objeto de estudo desta dissertação, foi iniciada nos anos 1956 com o artigo “Is Consciousness a Brain Process?” de U. T. Place, seguindo-se os artigos “‘The Mental’ and ‘The Physical’” de Herbert Feigl, 1958, e “Sensations and Brain Processes”, de 1959, de J. J. C. Smart. Outros filósofos também defenderam alguma forma de teoria da identidade, mas, para consideração, nos limitaremos em algumas obras de D. M. Armstrong e David Lewis. Com isso, atenção especial é dada ao que podemos chamar de materialismo australiano, que se refere aos textos desenvolvidos na Austrália por U. T. Place, J. J. C. Smart e David Armstrong. Essa limitação não se deve apenas a fatores geográficos, mas principalmente à estrutura assumida por essas teorias.

Uma descrição das interações dentro dessa comunidade filosófica no período em que a teoria da identidade estava sendo desenvolvida é encontrada em (Moravia, 1995). Segue-se a sua narração dos acontecimentos:

Em 1955 U. T. Place deu uma série de palestras na Universidade de Adelaide na qual ele apresentou uma solução de identidade para o MBP [problema corpo mente, do inglês mind-brain problem] (ou, mais precisamente, para uma parte razoavelmente restrita desse problema). As reações iniciais de filósofos como Smart e Martin, que estavam trabalhando em uma teoria da mente, não foram favoráveis. Suas abordagens do relacionamento de mente e corpo eram condicionadas nesse

tempo pela interpretação do mental e suas funções como puramente linguísticas, dada por Gilbert Ryle, e por noções conceituais. Assim, parecia ilegítimo identificar consciência (como fez Place, embora cuidadosamente) com alguma *coisa* chamada ‘mente’. A evolução da reflexão e do debate sobre o MBP avançou rapidamente. Em 1957, Smart ministrou um curso sobre Wittgenstein e Ryle em Princeton no qual ele argumentou em favor de posições substancialmente similares àquelas de Place. Em novembro do mesmo ano, palestrando em Cornell, ele explicitamente expôs e defendeu a tese de Place. (Moravia, 1995, pp. 61-62)

Em termos gerais, a teoria da identidade sustenta que “a mente é o cérebro”. Já foi desenvolvido no capítulo 3 o que esses filósofos consideram como sendo o físico, mas quais entidades ou fenômenos mentais são objeto de identificação por parte desses autores sofreu alterações com o tempo. Neste primeiro momento, será feita uma breve exposição sobre as classificações de objetos mentais feitas por esses filósofos.

Inicialmente, U. T. Place (1970, pp. 42-3) distinguiu os conceitos mentais em duas categorias: os conceitos cognitivos, como ‘conhecer’, ‘crer’, ‘entender’, ‘recordar’, e os conceitos volitivos como ‘querer’, ‘pretender’, para os quais uma explicação behaviorista em termos de disposições para se comportar fazia sentido. Além desses, há outros conceitos, como consciência, experiências, sensações e imagens mentais, que deixariam resíduos intratáveis para uma explicação behaviorista, tornando-se então os objetos de identificação da *nova* teoria.

Herbert Feigl, no seu artigo “The ‘mental’ and the ‘physical’”, além de fazer uma análise ampla dessas duas categorias de conceitos, também fez uma distinção entre as entidades mentais. Haveria os sentimentos brutos (*raw-feels*), que seriam os eventos e processos da experiência direta; além destes, haveria os atos intencionais, como “eventos e processos inconscientes e atos intencionais de percepção, introspecção consciente, expectativa, pensamento, crença, dúvida, desejo, volição, resolução, etc.” que, por serem “ocorrentes da percepção direta que são descritos introspectiva ou fenomenologicamente” como sentimentos brutos, têm estes sentimentos brutos que lhes estão associados como objeto de identificação (Feigl, 1958). Assim, a identificação desse ocorrente como sentimento bruto não é excluída dentre os objetos a serem identificados, embora a identificação neurofisiológica dos atos intencionais enquanto tais levem a erro categórico.

J. J. C. Smart, em seu artigo “Sensations and brain processes” também fará uma distinção entre conceitos cognitivos, para os quais ele continuará sendo behaviorista, e as sensações, que serão os objetos de identificação com os processos mentais. Esta seria praticamente a mesma posição adotada por Place.

D. M. Armstrong, no livro *A materialist theory of the mind*, p. 80, apontará como uma das distinções entre sua teoria e as inicialmente desenvolvidas por Place e Smart, cujos artigos são considerados como sendo exposições “clássicas” do materialismo de estado central, a identificação de *todos* os estados mentais com estados cerebrais. Tal posição foi posteriormente adotada por Smart.

Em seu texto, “The ‘Mental’ and the ‘Physical’”, Herbert Feigl, na seção V-D, faz a seguinte afirmação que se aproxima da extensão da identidade defendida por D. M. Armstrong:

“[...] a neurofisiologia do futuro (3000 A. D. ?) deve fornecer derivações dedutivas completas dos vários estados centrais cujos ψ -correlatos familiares são as sensações, percepções, pensamentos, crenças, desejos, vontades, emoções e sentimentos (conhecidos por *acquaintance* e descritos em linguagem fenomenal).” (Feigl, 1956, V-D)

Essa passagem aponta uma tendência de Feigl por uma identificação abrangente dos estados mentais e físicos, que precede a desenvolvida por Armstrong e depois aceita por Smart. Em outra passagem da mesma seção, ele afirma:

“Conceitos de teoria de comportamento molar como força de hábito, expectativa, *drive*, instinto, traço de memória, repressão, superego, etc., ainda podem ser identificadas em uma psicofisiologia futura com tipos específicos de estrutura-neural-e-padrões-de-processos.” (Feigl, 1956, V-D)

Assim, acompanhamos como o objeto de identificação sofreu alteração ao longo do desenvolvimento da teoria da identidade, inicialmente proposta por Place, até sua formulação na teoria de estado central de David Armstrong.

Um outro aspecto da teoria da identidade está vinculado com o reducionismo, porque muitas vezes a teoria da identidade é considerada como sendo reducionista e, por isso, uma forma de se opor a ela é negar a existência de estados mentais irreduzíveis. Alguns filósofos são mais reducionistas que outros. Por exemplo, Smart admite algumas vezes a redução ontológica, como foi exposto na seção 4.3 em que se tratou do fisicalismo. Dentre os temas desenvolvidos pela teoria da identidade, um deles deveria tratar da irreduzibilidade/reduzibilidade dos estados mentais a estados cerebrais. Apesar disso, será feita apenas uma breve exposição com relação à irreduzibilidade.

Segundo Feigl (1958), há dois sentidos para irreduzível. Em um primeiro sentido, pode

significar não derivável de um conjunto específico de premissas; e em um segundo sentido, pode significar não-tradutível, quer dizer, que não é sinônimo ou equivalente no sentido lógico. Smart (1959, 1962), por exemplo, deixará evidente que a identidade não é translacional ao afirmar segue-se que sua tese de identidade não sustenta que afirmações de sensação podem ser traduzidas em afirmações sobre processos cerebrais. O mesmo sendo afirmado por Armstrong (1993, p. 84), embora em um contexto diferente, ao falar sobre sua fórmula de definição de estados mentais: “não deveria ser considerada como um guia para a produção de traduções de afirmações mentais. Pode bem ser que não seja possível traduzir afirmações mentais em afirmações que não mencionem nada além de acontecimento físicos”. Como foi exposto anteriormente no capítulo 3, sobre fisicalismo, trata-se principalmente de uma redução ontológica.

Na filosofia da mente, a teoria da identidade identifica contingente e ontologicamente estados e processos mentais com estados e processos cerebrais. As palavras “contingente” e “ontologicamente” não entram em uma definição padrão dessa teoria, como a do artigo “The identity theory of the mind”, escrito por Smart para a *Stanford Enciclopeadia*, mas os argumentos que foram levantados continuamente em sua defesa sustentam e incentivam uma definição nesses termos. Além disso, essa é a linguagem utilizada para apresentar a teoria no livro de Smart *Nosso lugar no universo*. Mais uma vez, a ênfase no caráter ontológico da identificação decorre principalmente da justificação do reducionismo fisicalista sustentando por Smart.

Nessa identificação, conceitos mentais são inicialmente classificados como estados, processos ou eventos, como o faz Armstrong, e estes são posteriormente identificados com estados, processos ou eventos cerebrais. Essa distinção não é desprezível, pois ajuda a identificar conceitos que eram geralmente considerados puramente mentais e que não eram cogitados para identificação com estados ou processos cerebrais, como por exemplo os desejos e intenções. Estes conceitos eram considerados puramente mentais e definidos de acordo com o método behaviorista na primeira fase da teoria da identidade. Ou seja, tanto Place quanto Smart passaram a identificar percepções e sensações a estados e processos cerebrais, mas continuavam behavioristas com respeito aos estados mentais superiores como crenças, desejos e intenções. Só após o trabalho de Armstrong, Smart, pelo menos, ampliou a identidade para todos os estados mentais e não apenas para aqueles de sensação e percepção. Portanto, a identidade inicial de fenômenos mentais a estados mentais foi um movimento importante e, pode-se dizer, fundamental para a identificação da totalidade dos fenômenos

mentais aos estados e processos cerebrais.

Identificar a totalidade dos estados e processos mentais e cerebrais não é o mesmo que identificar a mente e o cérebro. As vezes em que Smart e Armstrong falaram em termos de identidade entre “mente” e “cérebro”, houve a ressalva de que esse modo de expressar não é rigoroso mas uma rude simplificação. Como exemplo de uma passagem na qual ele faz essa observação, pode-se citar o início do artigo “The identity theory of mind”, no qual Smart afirma: “Falando rigorosamente, não é preciso sustentar que a mente é idêntica ao cérebro [...] Aqui eu tomo identificar mente e cérebro como sendo uma matéria de identificar processos e talvez estados da mente e do cérebro.”²⁸ Também Armstrong escreverá na abertura do sexto capítulo de *A Materialist Theory of the Mind* (p. 73): “a mente é simplesmente o sistema nervoso central ou, menos acuradamente porém mais epigramaticamente, a mente é simplesmente o cérebro.” Portanto, não é rigoroso afirmar que essa teoria identifica o cérebro com a mente.

Por que começar distinguindo essas duas formas de identidade (identidade cérebro-mente e identidade de estados e processos mentais com estados e processos cerebrais)? Simplesmente porque alguns filósofos apontam a identidade em termos de mente e cérebro como sendo absurda. Diziam, por exemplo, que um defunto tem cérebro, mas não tem mente e tentavam, com isso, ridicularizar a teoria da identidade. Dessa forma, se não há um certo tipo de processo ocorrendo em um cérebro e ele está em um estado que pode ser caracterizado como inativo, então pode-se supor que certos processos mentais não ocorrem e, portanto, a mente está também inativa. Não exatamente com relação à mente, mas no que diz respeito às caracterizações do que é estar vivo ou morto, as análises feitas em hospitais para caracterizar a morte clínica parecem seguir uma via parecida com essa. Os óbitos são determinados pela falta de funcionamento do cérebro em um determinado nível, não com a presença ou ausência da massa encefálica. Supõe-se que quando o cérebro para de funcionar, também a mente para e não haverá processos mentais como dor, crenças, esperanças etc., caso contrário, considerações éticas poderiam impedir a tomada de procedimentos que pudessem causar dor no “paciente”.

Depois desta primeira apresentação da teoria da identidade, de como ela é formulada depois de todo desenvolvimento teórico, formulação essa que é apresentada nos materiais pesquisados que foram escritos e publicados nos anos 1990, darei ênfase às principais formulações dessa teoria ao longo do seu desenvolvimento, começando por Place.

²⁸ Texto sem paginação, artigo da internet, conforme referência.

No artigo “Is consciousness a brain process?”, de 1956, Place defende que a identidade entre consciência e processo no cérebro seja posta como uma hipótese científica razoável, que não pode ser tratada apenas como uma questão lógica, quer dizer, não é uma questão filosófica²⁹. Posteriormente, no artigo “Materialism as a scientific hypothesis”, originalmente publicado em 1960, essa divisão de tarefa é evidenciada, conforme as seguintes passagens:

O que é importante é que deve haver alguns critérios lógicos os quais usamos na decisão de se dois conjuntos de observações correlatas se referem ao mesmo evento ou a dois eventos separados mas causalmente relacionados. [...] Nós mesmos nos encontramos cara a cara com uma questão puramente empírica, nomeadamente, se há de fato um processo fisiológico, seja ele no cérebro, no coração, no fígado, no rim ou no dedão do pé, que satisfaz o critério lógico requerido para estabelecer sua identidade com o processo de sensação. (Place, 1970, p. 84)

Esse artigo foi uma resposta a Smart, por este ter afirmado que o problema mente-corpo, consiste de uma questão de decisão entre o materialismo de um lado e do outro as teorias paralelistas, epifenomenalistas, emergentistas, não pode ser decidido como uma questão empírica, mas em razão de plausibilidade e simplicidade.

Aparentemente, o critério de decisão é apontado no primeiro parágrafo do artigo de Place (1994 [1956]): a tese da identificação entre a consciência e um padrão de atividade cerebral tem como hipótese a “explicação das observações introspectivas do sujeito pela referência aos processos cerebrais com os quais eles estão correlacionados”. Esse critério pode ser formulado da seguinte forma:

Para n natural, considere os processos cerebrais $p = p_1, p_2, \dots, p_n$ de um sujeito x e as observações introspectivas $i = i_1, i_2, \dots, i_n$, tal que p_j esteja correlacionado com i_j , para j variando de 1 a n . Se existe uma explicação de i por meio exclusivamente de uma referência a p , então a consciência é igual ao padrão de atividade cerebral dado.

Note que, se a interpretação acima está correta, a conclusão é universal, mas os elementos da correlação e os testes para verificar a explicação das observações são particulares. A hipótese de Place não é sobre processos mentais individuais, como pode ser verificado em Place (1994 [1956]), onde a afirmação de identidade defendida é geral ou universal. Quanto a isso não há problemas, enquanto se toma uma visão falsificacionista da ciência, já que os testes podem servir para tentar falsificar a teoria, embora seja difícil dizer se

²⁹ A mesma opinião terá Armstrong quando propuser sua teoria, ou seja, uma divisão de tarefas entre a ciência e a filosofia, no estabelecimento da teoria da identidade.

ela é verdadeira como alguns filósofos desejam.

As observações introspectivas que fazem parte do conjunto a ser identificado são aquelas representadas pelos conceitos de consciência, imagem mental, experiência e sensação, sendo excluídos aqueles conceitos tidos como cognitivos, como conhecer, acreditar, entender etc., que são tratados por meio de comportamentos ou disposições para se comportar. São os eventos e processos (depois também estados) internos que este grupo de filósofos está preocupado em mostrar não só que existem, mas que são iguais aos eventos e processos cerebrais. Nesse fase, Place passou a sustentar simultaneamente a teoria da identidade para os processos e eventos internos e o behaviorismo para os conceitos cognitivos. Infelizmente eu não pesquisei suas opiniões posteriores, pois Place deixou de publicar em filosofia, mantendo-se na área da psicologia.

Em termos modais, a identidade é posta como não sendo necessariamente falsa, e também como não sendo necessariamente verdadeira, ou seja, por exclusão, é uma afirmação possivelmente verdadeira. Além disso, Place evidencia o caráter não apriorístico da identidade, ao negar à identidade o caráter de definição através do “é” presente na afirmação, e afirmar que ele tem um sentido de composição. Para fazer esta distinção, ele aponta os seguintes exemplos: para o “é” de definição, as afirmações: “Um quadrado é um retângulo equilátero”, “Vermelho é uma cor” e “Entender uma instrução é ser capaz de agir apropriadamente sob as condições apropriadas”; e para o “é” de composição, as afirmações: “A mesa dele é um caixote velho”, “O chapéu dela é um feixe de palha amarrado junto com um cordão” e “Uma nuvem é uma massa de gotículas de água ou outras partículas em suspensão”.

Uma característica dessa identidade é que Place afirma que, em ambos os casos, as afirmações informam tudo o que é preciso, fazendo sentido acrescentar a cláusula “e nada mais” no final de cada uma das afirmações. É possível verificar isso substituindo o “é” por “é composto por”, como em “O chapéu dela *é composto por* um feixe de palha amarrada junto com um cordão e nada mais”.

De acordo com os exemplos e explicações, pode-se afirmar que as definições se aplicam a todos os exemplares dos objetos que são definidos. Assim, todas as instâncias que são vermelhas são cores, todos os quadrados particulares são retângulos equiláteros e o mesmo se aplica para o entendimento de uma instrução. Observe que pelo menos duas das afirmações são ou podem ser interpretadas como universais e, sendo o quadrado o nome de uma figura abstrata, também pode ser interpretada dessa forma. As afirmações de composição, diferentemente, são afirmações particulares, exceto talvez a última. O “é” relaciona *situações*

que não são comuns a todos os objetos aos quais se referem. Um pouco de dificuldade poderia ser encontrado se fosse observado que a afirmação de identidade que Place está defendendo é uma afirmação geral ou universal e as afirmações utilizadas como exemplo e a defesa feita por meio delas têm como base afirmações particulares. Aparentemente a defesa de Place parece não funcionar para uma identidade tipo-tipo, já que a composição particular de uma mesa faz com que seja verdade que ela seja composta por um caixote velho, mesmo que não seja uma característica de todas as mesas. Em outras palavras, esse argumento parece defender que mesmo quando a afirmação universal é falsa, a afirmação particular pode ser verdadeira.

A justificativa dada por Place para esse problema e a forma como ele o soluciona faz uso da consideração da independência lógica dos termos que estão sendo identificados. Há uma regra que diz que quando duas expressões são logicamente independentes e se referem a um mesmo objeto ou estado de coisas, então “devem se referir a características ou um conjunto de características que não são normalmente ou necessariamente aplicadas ao objeto ou estado de coisas em questão” (Place, 1970 [1956], p. 25). Assim, normalmente se conclui que esses objetos ou estado de coisas são ontologicamente independentes. A solução encontrada foi considerar a identidade “A consciência é um processo no cérebro” como uma exceção à regra, devido à falta de “simultaneidade para se fazer as operações de verificação da presença dos dois conjuntos de características inerentes à coisa ou estado de coisa”, conforme ele afirmou na mesma página. Neste ponto são aplicados os testes empíricos que permitirão que a identidade seja verificada. Desse modo, fica claro o papel que Place atribui à experiência empírica na identificação da consciência com os processos do cérebro. Armstrong fará um desenvolvimento desse método, como será exposto no capítulo mais à frente.

Uma característica da teoria da identidade apresentada por Place, Feigl, Smart e Armstrong é questão da realidade dos estados mentais, das experiências mentais. Place faz uma distinção em um dos parágrafos finais do seu artigo de 1956 de que não se está afirmando a realidade de coisas como pós-imagens verdes³⁰, que existem e são verdes, mas de uma experiência que normalmente se tem quando se está diante de uma coisa real que seja verde e que se aprende a relatar dessa forma. Essa distinção também será adotada por Smart. Há uma distinção e analogia entre a experiência diante de um objeto real e a experiência sem a presença desse objeto. O que Smart e Feigl enfatizam: a realidade do que está sendo relatado

³⁰ Imagens obtidas, por exemplo, após olharmos fixadamente para um objeto vermelho, e depois voltarmos a olhar para uma parede branca.

e, no caso de Armstrong, a realidade do que está sendo definido, que os objetos mentais são reais. Um aspecto comum a praticamente todos esses filósofos é a distinção fregeana entre sentido e referência, fazendo com que todos eles façam a “identificação da referência”, ou seja, o objeto que está sendo referido é que será identificado, levando a uma identidade ontológica. Assim, Feigl afirma:

Quando o médico me pergunta se tenho uma dor no meu peito, se meu humor está sombrio, ou se posso ler as letras miúdas, ele pode se dar ao luxo de ser um behaviorista e testar essas várias experiências de uma maneira perfeitamente objetiva. Mas eu tenho (ou não tenho) a dor, o humor depressivo, ou a sensação visual; e posso relatá-los sob a base de experiência direta e introspecção.
(FEIGL, 1958, I)³¹

Feigl levanta a seguinte questão, apresentada no modo material do discurso do seguinte modo: “como os sentidos brutos estão relacionados com estados comportamentais (ou neurofisiológicos)?”; e, no modo formal do discurso: “quais são as relações lógicas do discurso de sentido-bruto [*raw-feel-talk*, em inglês] (termos fenomenais, se não linguagem fenomenal) com os termos e declarações na linguagem de comportamento (ou de neurofisiologia)?”

No caso da identidade defendida por Feigl, o que está sendo identificado é o referente de termos subjetivos com os referentes de termos objetivos. A abordagem de Feigl não pode deixar de receber o rótulo de uma análise lógica, certamente uma herança do Círculo de Viena, que ele cambiou do positivismo lógico para o menos radical empiricismo lógico. Inclusive, quando, assim como Place, ele faz a divisão de tarefas, deixa a ciência com a parte psíquico-fisiológica ao passo que, a parte lógica e epistemológica de clarificação de conceitos fica a cargo da filosofia.

Contra que teorias Feigl desenvolve a sua? Além do behaviorismo, ele aponta o emergentismo e o paralelismo, porque o que lhe interessa pesquisar são as leis de correlação, que o faz colocar em questionamento a causalidade entre mente e cérebro. Ele argumenta contra a teoria epifenomenalista ou emergentista, que além das relações causais normais entre as coisas físicas, acrescenta também leis “pingentes” (*dangling laws*) que regulariam as correspondências entre as coisas mentais e as físicas.

Uma proposta de solução para esse problema, que será utilizada mais tarde pelos filósofos da teoria causal da mente, é considerar a concomitância em termos das “estruturas

31 Uma passagem muito próxima também pode ser encontrada em Feigl (1970 [1960], p. 34). Novamente ele nega a abordagem behaviorista, o limite que essa teoria teria em falar da mente, e afirma: “but it is *myself* who *has* them.” É sempre interessante observar que os sujeitos *têm* estados mentais.

idênticas subjacentes” (Feigl, 1970, p. 36) Esse critério é derivado da ciência e consiste em explicar certas propriedades de alguns elementos em função da estrutura molecular subjacente.

Com relação ao status ontológico do que está sendo objeto de identificação, Feigl considera a identidade psicofisiológica tanto entre particulares quanto entre universais, ou seja, tanto identidade de uma instância mental com uma instância neurofisiológica quanto identidade de um tipo físico universal com um tipo mental.

Para concluir a apresentação da teoria de Feigl, será feito este adendo. A teoria de estado central apresentada por Armstrong se aproxima da que foi anteriormente apresentada por Feigl, pois há uma cadeia de identidades na qual os estados brutos de sentido são identificados indiretamente com os estados físicos do cérebro, além de haver uma semelhança nas terminologias. Feigl apresenta da seguinte maneira a sua teoria da identidade:

[...] o que é *tido-na-experiência*, e (no caso dos seres humanos) conhecido por *acquaintance*, é idêntico com o objeto do conhecimento por descrição providenciado pela teoria molar do comportamento e este é, por sua vez, idêntico com o que a ciência da neurofisiologia descreve (ou melhor, descreverá quando progresso suficiente tenha sido alcançado) como processos no sistema nervoso central, talvez especialmente no córtex cerebral.

(Feigl, “The ‘mental’ and the ‘physical’”, V-E)

Assim, como foi mencionado na introdução com relação à *especificação funcional*, espécie de funcionalismo assim classificado por Ned Block por identificar os estados mentais a estados funcionais e estes a estados físicos, a identidade defendida por Feigl apresenta claramente essa dupla identificação. Os termos utilizados por Feigl em seu artigo (por exemplo, “*drive*”, “força de hábito”, “teoria molar de comportamento”, usados na própria declaração da teoria da identidade) remetem ao behaviorismo, especificamente aquele da linha de Edward Tolman e Clark Hull. O primeiro psicólogo fez uma distinção entre os comportamentos com e sem propósito, chamando de molares aqueles padrões de comportamento que eram significativos, ou seja, que tinham um propósito ou eram direcionados a um fim; e de moleculares os movimentos musculares simples como os reflexos musculares, por exemplo. Para explicar esses comportamentos, inicialmente Tolman acrescentou entre os estímulos e as repostas comportamentais um mecanismo interno de “variáveis intervenientes”: tais variáveis seriam como mapas cognitivos que representavam o

ambiente com suas rotas, caminhos e relações ambientais. Nesse esquema, é esse mapa que vai determinar se e que comportamento um dado estímulo vai gerar. Tolman teria apresentado dessa forma sua teoria ainda em 1948, no artigo “Cognitive Maps in Rats and Men”.

Para prever o comportamentos a partir de um dado estímulo, Hull elaborou um esquema no qual a resposta comportamental (as variáveis de *output*) era implicada por uma conjunção de estímulo externo (que seriam as variáveis de *input*) e de estados internos do organismo (as variáveis intervenientes introduzidas por Tolman). Esta implicação é considerada como um postulado (suponha que seja o postulado *H*)³² a partir do qual outras afirmações sobre o comportamento e as relações entre as variáveis são deduzidas. Entre as variáveis intervenientes encontramos alguns dos termos que Feigl expressa algumas vezes em seu artigo, que são *drive* e força de hábito (*habit strength*). As relações entre essas variáveis e seus atributos seriam derivadas na forma de um sistema dedutivo a partir do postulado (H) acima.

A estrutura do postulado *H* aparece como uma espécie de prévia da estrutura que encontramos nas teorias funcionalistas e, o que é relevante para este estudo, foi uma estrutura que Feigl conhecia e que utilizou, de certa forma, na sua afirmação. Ao identificar os estados de sentimento brutos, ele está inserindo, embora não explicitamente, esses estados na estrutura apresentada por Hull. Veja que, se for considerada dessa forma, o enunciado de identidade pode se tornar equivalente ao apresentado por David Armstrong, que será apresentado mais adiante, e todo movimento que Armstrong faz para afirmar a identidade está presente implicitamente no texto de Feigl. Observe que se os estados brutos forem identificados com “o objeto do conhecimento por descrição providenciado pela teoria molar do comportamento”, então esse estado está sendo incluído na cadeia de relações entre os estímulos (*inputs*) e respostas comportamentais (*outputs*). Observe também que há primeiro uma identificação com esse objeto (variável interveniente), seguida de uma identificação com os processos no sistema nervoso central.

Dadas as teorias de Place e de Feigl, Smart partiu das teorias da identidade desenvolvidas por esses dois filósofos, portanto, como já foi apontado no texto, há um conjunto de características comuns às teorias desenvolvidas por esses filósofos. Na apresentação da teoria de Smart, será dada prioridade às poucas diferenças que podem ser identificadas.

Antes de prosseguir, gostaria que fosse observado que, segundo o que foi visto até este

³² Considerando a afirmação da implicação como um postulado.

ponto, e isso também acontece com Smart, não existe uma coisa chamada pós-imagem, não existe uma dor etc., o que existe é a experiência de uma pós-imagem, a experiência de uma dor etc. Nesses termos, fica mais difícil entender a identificação dos *objetos* mentais com os processos cerebrais, porque a dor como experiência fica mais abstrata e difícil de manipular. Quer dizer, é mais fácil substanciar um objeto mental e atribuir-lhe propriedades, que teorizar sobre a experiência e mesmo definir uma experiência.

Foi mostrado que tanto para Place quanto para Feigl a identidade deveria ser feita sobre dois conjuntos de coisas *correlacionadas*, processos e eventos mentais, de um lado, e processo e eventos cerebrais de outro. O primeiro ponto de distinção de Smart é negar a correlação e portanto a necessidade de leis que expressem essa correlação entre os objetos desses dois conjuntos, baseado no fato de que um objeto não pode ser correlacionado com ele mesmo, e a correlação diria que os processos mentais estariam além dos processos físicos. Aparentemente, esse aspecto implicaria na forma metodológica da identidade, pois atingiria a forma como os objetos passíveis de identificação seriam considerados.

A crítica sobre a correlação consiste em perceber que tal correlação admite que os objetos i_j do conjunto de observações internas sejam de uma natureza diferente da dos objetos p_j do conjunto dos processos cerebrais, o que vai contra o espírito da teoria da identidade. Algumas das críticas à teoria da identidade recaem justamente sobre esse ponto, porque alguns filósofos admitem que todo relato de observação interna é um relato de algo que não é físico e então se colocam as questões de como fazer a identificação, havendo inclusive a sugestão de leis psicofisiológicas relacionando as observações mentais “não físicas” de um lado e os processos cerebrais de outro.

A solução encontrada por Smart para resolver o problema dos relatos de experiências mentais foi apresentada como resposta à Objeção 3 no artigo “Sensations and brain processes”. Essa solução tem um papel principal na presente dissertação, que pesquisa a relação entre teoria da identidade e funcionalismo. Neste ponto, há uma interação entre os três filósofos, Smart, Armstrong e Lewis, para o seu desenvolvimento, tendo originado a suposta confusão entre funcionalismo e teoria da identidade da qual o capítulo 2 fez uma exposição.

A resposta à Objeção 3 é justamente uma forma de se esquivar do relato dos estados internos carregados de propriedades não físicas. Esta objeção afirma que mesmo se for provado que as sensações não são nada mais que processos cerebrais, as qualidades das sensações continuam sendo qualidades não físicas. Este é o famoso argumento dos *qualia* que ainda surge na filosofia da mente.

Antes de apresentar a solução de Smart, é preciso fazer uma elucidação do objeto

mental que é o protagonista principal desta história. Existem as sensações que estão relacionadas com os objetos que estão diante da pessoa que sente. Se estou vendo um pote verde sobre minha mesa, então eu sinto a verdura, tenho uma sensação que corresponde a esse objeto. Mas pode acontecer de anoitecer, eu acordar à noite no escuro total e ter uma visão de um objeto verde. Não há, de fato, nenhum objeto externo que cause em mim essa sensação, mas não posso negar que tenho uma imagem verde, que experiencio essa imagem. Um *quale* pode ser tanto a qualidade da minha sensação relacionada à verdura do objeto que vejo à minha frente, quanto a sensação também de verdura que tenho quando imagino, tenho a imagem mental de um objeto verde inexistente. Os *qualia* estão associados com a experiência fenomenológica de cada sensação.

Para solucionar o problema dos *qualia*, Smart introduz um artifício, na verdade, ele vai aplicar uma fórmula que está sob os critérios de uma identidade contingente, ou seja, aquele na qual os termos que estão sendo identificados são logicamente independentes. O argumento segue a seguinte ordem: (1) introdução do conceito de *percipiente normal*, que é erigido sobre a capacidade de se *fazer discriminações*³³. Assim, se uma pessoa consegue fazer mais discriminações de cores que outra pessoa, ela é mais normal que a outra. A capacidade de discriminar é feita por meio de testes que dificultam gradualmente a discriminação, como visualizar um objeto entre objetos com cores parecidas, por exemplo. Com isso, dizer “isto é vermelho” significa algo como “Um ser humano normal não poderia tirar isto facilmente de um porção de pétalas de gerânio, embora possa retirá-lo facilmente de uma porção de folhas de alface” (Smart, 1959, p.149). Smart acrescenta o fato de uma pessoa ser treinada para fazer essas discriminações, mesmo sendo uma pessoa cega que utilize para isso uma pessoa que possua visão normal. (2) Na mesma página ele apresenta a definição de cor como “o poder de evocar certas sortes de respostas discriminatórias em seres humanos.” As cores fazem parte das qualidades secundárias, na classificação lockeana, assunto ao qual Smart se dedicou por longo período.

(3) O próximo passo é definir a forma de relato da experiência interna. Assim, ainda na mesma página, Smart expressa:

Quando uma pessoa diz, “vejo uma pós imagem laranja amarelada”, ele está dizendo algo como isto: “*Há alguma coisa acontecendo em mim que é como o que acontece em mim quando tenho meus olhos abertos, estou desperto e há uma laranja iluminada em boa luz em minha frente, isto é, quando eu realmente vejo uma laranja.*” (Smart, 1959, p.149)

33 Em Smart (1959b, p. 40), *percipiente normal* é definido como um não *percipiente anormal*, sendo que um *percipiente anormal* não consegue fazer discriminações que um *percipiente* consegue.

Essa fórmula será usada para definir tanto os casos nos quais há um estímulo real no ambiente, como um objeto que causa as sensações relatadas na pessoa, quanto os casos em que há apenas uma imagem mental.

Observe que as experiências são relatadas como “*há alguma coisa acontecendo em mim que é como o que acontece em mim quando ...*”, sem que sejam expressos os nomes de quaisquer estados ou processos mentais. Na verdade, Smart explica que essa fórmula é tópico neutra, ou seja, ela não expressa nada sobre a natureza do que está sendo relatado. Não diz se a experiência é mental e também não diz se ela é física. Esta fórmula é neutra quanto ao dualismo e ao materialismo. Ela estabelece apenas uma relação de similaridade entre as experiências de uma pessoa. Esta fórmula acabaria com o problema de buscar uma lei que relacionasse os processos mentais aos processos físicos, já que ela é neutra e portanto não está relatando nenhuma propriedade não física ou física que precisa ser relacionada. Dessa forma, há uma única natureza ou objeto a ser pesquisado, a experiência subjacente que, por hipótese da teoria da identidade, é física.

Nesse artigo, Smart também aponta três explicações que essa forma de relato propicia: a neutralidade do camponês grego entre o dualismo e o materialismo, quando faz o relato de suas sensações; como as sensações podem ser processos cerebrais e esse fato ser ignorado; e a elusividade dos estados brutos, a inefabilidade dos *qualia*.

Gostaria de salientar a importância dessa posição adotada por Smart na relação que irá se estabelecer entre o desenvolvimento da teoria da identidade e do funcionalismo. Nos trabalhos subsequentes, há um diálogo entre Armstrong, Lewis e Smart, principalmente com relação ao estabelecimento de uma fórmula que melhor expressasse o relato de experiência.

Existem três textos principais nos quais Smart discute e apresenta a evolução de sua fórmula tópico neutra: “Reports of immediate experiences”, de 1971, “On some criticisms of a physicalist theory of colors”, de 1975, e “Consciousness and awareness”, de 2004. Principalmente nos dois primeiros textos, Smart discute as implicações de cada uma das versões da fórmula e apresenta as influências que teve no seu desenvolvimento. Entre as pessoas que mais o influenciaram encontram-se Armstrong e Lewis. Essas influências aproximaram ainda mais a teoria de Smart a uma espécie de funcionalismo, por causa da abstração empregada na definição das experiências e qualidades secundárias dos objetos.

6 Funcionalismo

A forma básica como uma teoria funcionalista define um estado mental é por meio de uma estrutura relacional composta por três classes de elementos: *entrada* (*input*) – recebimento de estímulos do meio externo; *estados internos* – incluindo relações entre estes estados; e *respostas* comportamentais (*output*). Além disso, como se pode perceber, nessa descrição se exclui qualquer referência às bases materiais da coisa que realizará esse estado.

A emergência do funcionalismo coincide com o crescente interesse e desenvolvimento da computação. Talvez possamos dizer que, embora uma forma de funcionalismo possa ser encontrado já em Aristóteles (ver Shields, 1990), o funcionalismo contemporâneo é fortemente influenciado pela teoria computacional e o desenvolvimento da inteligência artificial. Uma das principais teses do funcionalismo, a possibilidade de múltipla realização do mental, parece comprovar essa influência.

Definir funcionalmente um estado mental não é muito distinto de construir um algoritmo que pode ser implementado em um número variado de tipos de máquinas ou computadores. Um algoritmo é distinto do suporte físico que o realizará. Uma computação pode ser feita num microcomputador como os que temos em casa, numa máquina analógica como a de Babbage, com lápis e papel, com alguns riscos no chão e algumas pedrinhas, uma máquina hidráulica, um computador quântico (ainda teoricamente), no cérebro (para os que defendem a teoria de que o cérebro é como um computador), utilizando-se a população da China, etc.

Os computadores atuais, na verdade, são implementações finitas de máquinas bem abstratas, a máquina universal de Turing, como consta no capítulo 8 desta dissertação. John Heil, por exemplo, enfatiza o papel da abstração na determinação de uma máquina de computação, e que quando consideramos algo como tal, abstraímos sua composição física para considerá-la como um aparelho computacional (Heil, 2004, pp. 91-92).

Agora considere uma teoria funcionalista que identifica um estado mental com uma função. Sabemos que funções são “entidades” abstratas. Assim, pergunta-se: Existem objetos abstratos? Pode-se dizer que um triângulo é uma abstração de um desenho à lápis vermelho em uma folha de papel, ou de qualquer outra figura concreta de um triângulo? Ou o triângulo é um objeto puramente abstrato no sentido de ser aquilo que sua definição lhe confere e que suas construções físicas (desenhos no papel, no quadro, estruturas construídas de madeira, recortes de plástico, etc.) são como representações dessa entidade abstrata? Haveria um

mundo dividido entre *entidades abstratas* e *entidades concretas*?

Essa questão metafísica está intimamente relacionada com o funcionalismo. As definições funcionais estão em que relações com o mundo físico? Elas são entidades abstratas ou são simplesmente abstrações da forma como um ser humano percebe e interage com o ambiente? É procedente comparar um processo ou estado mental a uma computação, ou seja, eles têm a mesma natureza?

Podemos identificar dois comprometimentos do funcionalismo, uma explicação funcional e uma ontologia funcional.

A explicação funcional está relacionada com a engenharia reversa. Engenharia reversa se refere à atividade investigativa do funcionamento de um equipamento do qual não sabemos ainda como funciona, nem como é composto e estruturado. Assim, por exemplo, um engenheiro, ou mesmo um curioso qualquer, se depara com um aparelho de ponta de um concorrente e tenta descobrir como ele funciona e do que é composto. A explicação funcional é justamente o relato de quais são as partes relevantes, como elas se relacionam, qual a função que cada parte exerce, como elas se combinam para o comportamento global do sistema, etc. No caso de humanos, tem-se a ideia de uma caixa preta, à qual se aplicam certos estímulos e tem-se acesso a certas respostas, como alguns experimentos realizados por cientistas da cognição e neurocientistas, ficando a cargo de quem analisa criar as hipóteses sobre os estados internos que vinculam os estímulos às respostas. A explicação funcional considera a distinção de níveis de explicação, o nível físico e um nível mais elevado, o mental. Essa distinção tem um análogo na computação que é aquele entre o hardware e o software.

A ontologia funcionalista faz referência à distinção de níveis ontológicos, mas sustenta que esses níveis possuem propriedades diferentes e não são identificáveis. Enquanto no primeiro caso a função é mais como uma ferramenta para tentar adivinhar como a máquina, ser humano ou o que estiver sendo estudado funcionam, neste caso a estrutura funcional faz parte da ontologia do mundo e, no caso do funcionalismo na filosofia da mente, pode ser identificado com a mente.

O que nos interessa realmente no funcionalismo é sua estrutura de explicação. A estrutura de definição funcional permite a múltipla realização do mental, e é sobre essa característica que faremos a comparação entre a teoria da identidade e o funcionalismo, levando-se em conta que a teoria da identidade tem essa característica em função especialmente de sua neutralidade.

6.1 Teoria Causal da Mente

No seção anterior, o funcionalismo recebeu uma abordagem mais geral. Agora, as teorias causais desenvolvidas por David Armstrong e Lewis serão consideradas mais detidamente, tendo em vista suas relações com a teoria da identidade. Na estrutura atual da dissertação, os dois filósofos e suas teorias são considerados isoladamente, e depois terão suas teorias comparadas. Esta dissertação seguirá a classificação de Ned Block e a própria admissão de David Armstrong e tratará a teoria causal da mente como sendo a forma causal do funcionalismo.

6.1.1 Teoria Causal da Mente, de David Armstrong

Neste subtópico, será apresentada a teoria desenvolvida por David Armstrong. Ela foi apresentada com o nome de *teoria de estado central* na primeira edição, 1968, do livro *A Materialist Theory of the Mind* (Armstrong, 1993), para salientar a identidade de estados mentais com estados do sistema nervoso central. Na segunda edição do mesmo livro, o filósofo apresenta algumas outras denominações pelas quais a teoria passou a ser conhecida, assim, de *teoria causal*, que ele passou a usar no artigo “The causal theory of the mind” de 1977, avisa que passou a denominá-la também de versão causal do Funcionalismo. Além dessas, uma outra denominação sustentada para a teoria é de *materialismo de estado central*, pelo mesmo motivo relacionado ao nome teoria de estado central.

Em Armstrong (1993), a teoria é desenvolvida em dois passos, sendo o primeiro passo uma análise lógica ou conceitual dos conceitos mentais; e o segundo passo diz respeito à questão do que desempenha o papel causal resultante do primeiro, ou seja, o que é causa ou é causado da forma especificada (a hipótese da identidade dos estados e processos mentais com os estados e processos cerebrais). É neste passo que se encontra o que pode e foi chamado por algumas pessoas de teoria da identidade, e no livro foi expressa num capítulo de apenas 12 páginas intitulado “Identificação do mental com o físico”. São apresentados três motivos para a curta extensão do capítulo: o primeiro diz que isso se deve a que esse assunto “empírico” foi abordado na primeira parte do livro; o segundo, de ele já ter sido defendido por outros autores, tendo como exemplo J. J. C. Smart no seu livro *Philosophy and Scientific Realism*; e

o terceiro motivo, à suposição de que se a parte dois do livro, que se dedicou à análise dos conceitos mentais, foi bem sucedida, então a identidade contingente entre o mental e o físico não pareceria particularmente paradoxal ou audaciosa. Não é demais observar que essa divisão foi apresentada anteriormente pelos filósofos da teoria da identidade, uma parte lógica que seria atribuída à filosofia e uma parte empírica que seria deixada a cabo da ciência.

A teoria causal da mente é uma teoria analítica, no sentido de se ocupar com a análise dos significados de termos e conceitos mentais em seus aspectos causais, dizendo respeito a estados mentais serem tanto causas quanto efeitos. Como Armstrong expõe em (1977), essa teoria é vista como um meio de o filósofo não apenas contribuir para as grandes questões gerais, mas também para tratar das próprias coisas, ou seja, a análise conceitual não é vista como um fim em si mesma, mas um meio de se buscar conhecimento sobre o mundo. Neste caso, pode também ser entendida como uma geradora de hipótese científica. Além disso, esse papel da teoria é apontado por Armstrong como sendo importante para a identificação de estados mentais com estados físicos, como podemos ver na seguinte passagem:

Meu ponto de vista é que a identificação de estados mentais com estados físicos do cérebro é uma identificação perfeitamente inteligível, e que isso se torna claro uma vez que nós alcancemos um ponto de vista correto da análise dos conceitos mentais. Eu admito que minha análise dos conceitos mentais foi ela mesma adotada porque permite esta identificação, mas tal procedimento é um lugar comum na construção de teorias, e perfeitamente legítimo. (Armstrong, 1977)

Pode-se considerar então, que a análise conceitual é a principal característica da teoria de estado central. Essa análise não está inserida no programa funcionalista, mas no programa da teoria da identidade. Armstrong buscou por meio da análise conceitual desenvolver uma forma ou fórmula de definição que permitisse identificar contingentemente estados mentais a estados cerebrais, prosseguindo e colaborando especialmente com o trabalho de Smart para construir uma fórmula que fosse neutra. É importante observar que tanto a teoria de Smart quanto a de Armstrong não implicam a verdade da teoria da identidade, pois a questão da verdade é empírica. Toda análise lógica consiste em possibilitar uma neutralidade com relação especialmente aos estados mentais, por isso esses filósofos buscam uma fórmula que elimine todo conceito mental das definições ou relatos de experiência. Conseguir definir uma teoria em termos neutros permite facilitar a identidade ontológica, já que haverá menos entidades a serem identificadas e leis que devam ser criadas para relacionar essas entidades. Lewis também procederá da mesma forma, ou seja, eliminando os termos mentais, mas,

diferentemente destes filósofos, desenvolverá uma prova lógica da identidade. Com isso, a verdade da teoria da identidade, com Lewis, deixa de ser um problema apenas empírico.

Antes de prosseguir para questões mais pontuais da teoria causal da mente, serão retomados os pontos levantados por Armstrong nos tópicos sobre o dualismo e o behaviorismo que toda teoria da mente que fosse boa deveria dar conta. A teoria do estado central, segunda a interpretação do próprio Armstrong, não dará conta de todos os requisitos que foram levantados.

A teoria do estado central é definida como “o ponto de vista que identifica estados mentais com estados puramente físicos do sistema nervoso central” (Armstrong, 1993, p. 73). Para possibilitar a identificação contingente entre esses dois estados, eles deveriam receber definições que os tornassem independentes um do outro. Assim, a mente deve receber uma definição na qual os o conceito de cérebro não está incluído, e vice-versa. Um outro termo utilizado por Armstrong é definição ostensiva: mente e cérebro devem ser ostensivamente definidos para que a identidade entre mente e cérebro possa ser feita contingentemente. Além disso, essa definição levaria a uma neutralidade com relação à natureza dos estados mentais, não se afirmando que são estados puramente psíquicos ou físicos, já que a identidade com os estados físicos será questão de pesquisa empírica científica.

O termo “cérebro” é definido como “um certo tipo de objeto físico encontrado dentro dos crânios de pessoas” (Armstrong, 1993, p. 78). Inicialmente, esse conceito é definido em termos behavioristas, quer dizer, utilizando termos behavioristas, na seguinte forma: “o efeito dentro de um homem de certos estímulos, e a causa dentro de um homem de certas respostas” (Armstrong, 1993, p. 79). Em seguida, o conceito de estado mental recebe duas versões, uma primária, definida como “*um estado da pessoa apto para causar um certo gênero de comportamento*”, e, para alguns estados de forma adicional, “*estados da pessoa aptos para serem causados por um certo gênero de estímulos*” (Armstrong, 1993, p. 82). Observe que em nenhuma das definições há vínculo literais com estados psíquicos.

Uma diferenciação com relação às teorias de Place e Smart, na teoria desenvolvida por Armstrong, está no fato de ele passar a identificar todos os estados mentais com estados cerebrais e não mais apenas os sentidos. Outra distinção está na definição em termos de comportamentos e disposições para se comportar e não apenas de estímulos. Observe que, embora Armstrong faça essa distinção, ela se deve à inclusão de todos os estados mentais na identificação, tendo os sentidos também uma definição com ênfase nos estímulos que atuam sobre eles, vide a forma adicional acima.

Com relação aos termos da definição, a palavra *estado* não está aplicada em sentido

rigoroso, mas guarda lugar para processos ou eventos também, se o conceito a ser definido não for um estado mas um processo ou evento. A palavra *apto* permite que estados mentais ativos que não resultam em comportamento sejam também definidos; além disso, engloba os estados mentais que não estão relacionados diretamente com comportamentos, mas fazem parte de uma cadeia causal de estados que estão relacionados em última instância com um comportamento. A causalidade envolvida é o conceito ordinário de causalidade eficiente, a causalidade científica. Quando ao conceito de comportamento, ele aparece na fórmula apenas como comportamento físico, para evitar circularidade. Tendo feito isso, convém analisar brevemente a forma como Armstrong analisa sua teoria com relação aos pontos solicitados às teorias da mente.

Observe que os problemas a seguir se aplicam quando se incorpora a identidade empírica, e não estão relacionados diretamente com a parte conceitual da teoria de estado central. Pode-se, no caso, fazer uma distinção entre os dois momentos de construção da teoria e analisá-los separadamente quando à satisfação desses pontos. Armstrong não faz isso, ele aborda a teoria de estado central junto com sua identificação empírica. Se for feita uma análise apenas da parte lógica, o resultado será diferente.

O problema (i), que diz respeito à possibilidade lógica de existência da mente desencorpada, segundo Armstrong, não é satisfeito pela teoria da identidade, já que a mente é definida como *estado da pessoa*. Apesar disso, *homem* é definido pela psicologia em termos funcionais ou behavioristas, “Homem é um objeto sobre o qual certos estímulos físicos agem continuamente” (Armstrong, 1993, p. 78). Esse problema pode ser respondido adequadamente de acordo com a forma como o termo “pessoa” é definido dentro de alguma teoria da mente.

Com relação ao ponto (ii), que trata da impossibilidade de se tratar os estados mentais como substâncias, Armstrong argumenta que a teoria de estado central também não satisfaz esse ponto, porque os processos cerebrais podem ser vistos como padrões de disparo elétrico no espaço e, enquanto tal, podem existir independentemente.

Quanto ao (viii) ponto, que trata da intencionalidade, na época em que o livro *A materialist theory of the mind* estava foi escrito, este era um ponto problemático porque Armstrong não estava seguro de poder respondê-lo adequadamente. Posteriormente, ele encontra uma solução para este problema, que é fazer uma analogia da intencionalidade ou consciência, se elas podem ser consideradas deste modo, com a propriocepção, que é o sexto sentido do corpo. Por meio do sentido proprioceptivo, somos capazes de saber a localização das partes do corpo sem necessidade de utilizar os outros sentidos para isso. A intencionalidade seria, então, uma percepção dos próprios estados mentais, uma consciência

da consciência. Como se uma parte do cérebro tivesse percepção de outra parte, gerando uma cadeia de estados perceptivos no qual o último não seria percebido, dando uma sensação de inefabilidade característica de certos níveis de consciência. Esta ideia é bastante aceita, especialmente por Smart, que a utiliza em artigos posteriores, especialmente no artigo (Smart, 2004). Além disso, também explica o fato de algumas vezes haver consciência de coisas no mundo sem que haja consciência dessa consciência, como acontece com o exemplo dado por Armstrong de um homem que dirige em uma estrada e não está consciente de dirigir, embora esteja consciente das coisas no mundo, caso contrário seria improvável que mantivesse o automóvel no caminho correto ou possível que se envolvesse em algum acidente.

Quanto aos demais pontos, eles são claramente satisfeitos.

6.1.1.1 O papel das disposições

No behaviorismo, os conceitos mentais eram explicados em termos de comportamento ou como disposições para um determinado comportamento. Uma crença era explicada como uma disposição para realizar um determinado comportamento, da mesma forma os desejos, etc. Armstrong toma as disposições e as aplica aos estados mentais como uma propriedade. Assim, a partir da sequência behaviorista:

estímulo → estado mental como disposição → comportamento,

Armstrong dará origem à seguinte sequência:

estímulo → estado mental → disposição → comportamento,

ou seja, o estado mental não será igual à disposição, mas será o estado que tem a disposição de causar determinado comportamento, neste caso, ou também terá a disposição de ser causado por um estímulo. Posteriormente, esse estado mental é identificado com um estado cerebral, um estado físico que possui essa disposição.

No livro *Consciousness & Causality* (Armstrong, 1984), o autor expõe uma teoria causal de disposições. Disposições são propriedades de objetos materiais como fragilidade e solubilidade, que influenciam ou determinam a forma como um objeto se “comporta”. Assim,

a fragilidade de um objeto o disporá a se quebrar ou rachar quando cair no chão cimentado; e a solubilidade do açúcar o disporá a se dissolver quando adicionado a um copo com água, por exemplo.

Armstrong dirá que um termo condicional que diz respeito à manifestação de certas propriedades de um objeto, por exemplo, a elasticidade de um elástico, depende causalmente da natureza do objeto e da aplicação de uma força. Em outras palavras, para ocasionar a manifestação da disposição, há uma relação entre o objeto e o mundo. Para mostrar que o elástico possui elasticidade é preciso que se aplique uma força a ele, embora seja evidente que o elástico pode nunca ser esticado e, portanto, ter evidenciada essa propriedade. Além disso, a elasticidade é atribuída ao elástico não apenas pela manifestação dessa propriedade (disposição), mas também devida a sua constituição interna, ou seja, de que elementos ele é composto e como esses elementos estão relacionados entre si.

Este aspecto é interessante para o presente trabalho porque essa dependência de uma natureza interna vai ultrapassar a explicação que se limitaria apenas a uma descrição condicional da forma: “se uma força é aplicada a um fio elástico de comprimento x , ele apresentará uma dilatação e seu comprimento será de $x + y$, após a remoção da força, o comprimento do fio voltará a ser x ”. Assim, a elasticidade de um objeto pode ser atribuída a sua constituição molecular, não sendo necessário iniciar um processo causal, como aplicar uma força para que ele se estique. Desse modo, Armstrong defende, em seguida, que se pode *identificar* essa disposição com a natureza física do objeto. David Lewis também utiliza esse ponto de vista com relação à disposição, uma tentativa de explicá-la em função da constituição interna dos elementos.

Com esse movimento, ele relaciona seu “funcionalismo” à teoria da identidade, já que ele parte de uma definição em termos das relações causais sobre um objeto para uma que leva em conta sua constituição interna, assim, relacionando uma explicação funcionalista com uma de identidade física. Mesmo que a constituição interna do objeto possa ser considerada uma causa de seu comportamento, deve-se observar que o fato de se tomar a natureza material do objeto na explicação constitui-se um elemento que distingue a metodologia entre uma teoria funcionalista e uma teoria da identidade. No primeiro caso, não importa a constituição material do objeto, se ele apresentar o comportamento de esticamento com a aplicação de uma força e depois voltar ao comprimento inicial quando essa força é removida, então ele possui a propriedade de ser elástico. No segundo, a constituição molecular do objeto será relevante para a explicação. Mas se se considera o funcionalismo como um está anterior na determinação do estado a ser objeto de investigação e identificação, então ele é indicado, pois

diminui a quantidade de entidades a ser identificada e, com isso, diminui os problemas inerentes a uma identificação entre coisas que são postas como tendo naturezas distintas. Este aspecto das teorias com foco na neutralidade das descrições foi apontado anteriormente.

Para Armstrong, é um problema do behaviorismo a definição de uma propriedade como a relação entre uma disposição e a apresentação da disposição sem consideração das causas internas. Em outros termos, poderia-se dizer que o elástico estica porque ele tem a propriedade da elasticidade, tem essa disposição. Assim, “*se* algo age sobre o elástico de uma certa forma, *então* ele manifestará certos efeitos.” Mas, para ele, esse tipo de explicação não é suficiente.

Armstrong defenderá que para esse condicional ser verdadeiro, deve haver algo no mundo concernente ao objeto que faça esse condicional ser verdadeiro e que, segundo a ciência, esse algo é alguma característica ou características da microestrutura do objeto. É importante salientar que essa microestrutura não é apenas considerada por ele como parte da estrutura causal da disposição, mas pode ser identificada com ela. Ele utiliza as seguintes palavras, com *itálico* original no seu texto:

[...] nós podemos identificar a disposição, e algumas vezes identificamos, [...] com esse aspecto micro-estrutural da fita elástica. Não é linguisticamente impróprio dizer que a elasticidade da fita *é* essa micro-estrutura da fita. [...] Mas nós também podemos tomar a elasticidade da fita como sendo aquela(s) característica(s) da fita que é (são) *responsável* (eis), causalmente responsável, pelo fato de, se uma força suave for aplicada, então... (Armstrong, 1984, p. 139).

Mas a ciência pode não saber a causa “interna” de alguma propriedade apresentada por um objeto. Neste caso, a propriedade poderia ser considerada como aquilo que desempenha um determinado *papel causal*. E não é só isso, mesmo sabendo qual é a microestrutura que desempenha o papel causal da propriedade em um dado objeto, ainda há o problema de se definir a propriedade de modo geral a fim de incluir na mesma definição as várias microestruturas que são a causa dessa propriedade em objetos diferentes. Com isso, quer-se dizer que identificar uma propriedade com a microestrutura de um objeto dado, como identificar a elasticidade com a microestrutura do elástico, não apresentaria problema, pois parece claro que a elasticidade do elástico depende de sua microestrutura, é a sua microestrutura [embora esse passo, do “*depender*” para o “*ser*”, precise de um pouco mais de fundamento]. Por outro lado, dizer que a elasticidade em geral, ou elasticidade *simpliciter*, como Armstrong fala, é a microestrutura do elástico, fica mais complicado.

Então pergunta-se: qual é o problema que Armstrong quer responder? Ele quer uma propriedade que seja aplicada de modo geral mas que desempenhe um papel causal nos objetos dotados com essa propriedade. No caso de se definir a propriedade muito abstratamente, ou seja, como uma propriedade de segunda ordem (uma propriedade que é atribuída a outra propriedade), no final, ela não será causalmente responsável pela manifestação da propriedade em questão. Seria preferível encontrar uma definição que, para cada instanciação sua e por mais que as coisas no mundo sejam distintas umas das outras, por exemplo, com relação às suas microestruturas, possa-se apontar a coisa no mundo que seja a responsável causal da manifestação da propriedade.

Nesse cenário, definir uma propriedade em termos do seu papel causal seria uma descrição suficientemente abstrata que seria identificada, sem maiores problemas, com aquelas coisas físicas que desempenham esse papel causal em um objeto específico. Assim, a propriedade *elasticidade* pode ser definida em termos de seu papel causal. Diz-se que um objeto tem essa propriedade quando, tendo tamanho inicial estável, ao ser submetido a uma força de certa intensidade, o objeto adquire um tamanho maior que o inicial e, ao se interromper a força que lhe foi aplicada, ele retorna ao tamanho inicial. Tomando-se vários objetos diferentes, como um fio de silicone, uma fita a base de elastômero, uma mola, o brinquedo Stretch Armstrong³⁴, uma bolinha de borracha, constata-se que a cada um deles pode-se aplicar uma força e observar o resultado esperado segundo a definição de elasticidade acima. Eles apresentarão um comportamento quase perfeito de um corpo que possui a propriedade de elasticidade, visto que foi observado o papel causal dessa propriedade em cada um desses objetos. Sendo assim, busca-se em cada um deles a causa microscópica, molecular, que desempenha o papel causal em questão, a elasticidade. Observe que em cada um desses objetos a elasticidade pode ser determinada por uma causa material que não precisa ser necessariamente igual à causa material de outro objeto. Esse aspecto da teoria leva à múltipla realização dessa propriedade.

Já foi expresso neste trabalho que os filósofos Smart, Lewis, Armstrong e Place consideram a teoria da identidade como uma hipótese científica. Do mesmo modo, Armstrong expressa que a teoria causal da mente pode ser considerada como uma teoria científica na qual os estados mentais são entidades teóricas (Armstrong, 1984, p. 141). Em tal abordagem ele apresenta uma vantagem e uma desvantagem, que não são tão evidentes quanto o filósofo faz

³⁴ Um boneco musculoso elástico que pode ser esticado, assumindo uma dimensão bem maior que a inicial. Ele fez bastante sucesso nos Estados Unidos em uma década do século passado.

supor.

A primeira diz respeito ao problema das outras mentes, ou seja, temos conhecimentos de nossa própria mente, de nosso próprio corpo e de algumas relações, principalmente causais, entre eles. Quando observamos outros corpos, seus atos e comportamentos, e quando comparando-os com os nossos percebemos que agem da mesma forma quando são submetidos aos mesmos estímulos e situações, supomos que estes também possuem os estados internos que nós possuímos. Embora seja aceita a tese de que não sabemos quais são esses estados, o que esses filósofos pretendem com as fórmulas tópico neutras é mostrar que não sabemos quais são esses estados. Ele afirma que esse tipo de hipótese possui uma fraqueza que é a extensão da base epistemológica, pois se faz a inferência a partir de um único caso, o do próprio indivíduo.

Para este problema, a teoria causal apresenta a vantagem de não privilegiar o acesso de primeira pessoa. Ambos os casos podem ser considerados a partir do comportamento atual na presença de determinados estímulos. A base para isso é que os seres humanos costumam apresentar uma regularidade de comportamento, especialmente sob o mesmo tipo de estímulo. E é a partir desse fato que a mente pode ser postulada. Este método evidencia a relação que essa teoria tem com o behaviorismo, pois a justificação para a postulação da mente como uma entidade teórica é comportamental, embora o método adotado por Armstrong seja analítico.

O filósofo aponta para o fato de que a desvantagem decorre justamente da vantagem, ou seja, de se considerar a mente do próprio sujeito como sendo um postulado, pois afirma-se que mesmo se não há ciência direta da mente das outras pessoas, parece não haver dúvida sobre a ciência direta dos próprios estados mentais.

Armstrong lida com essa desvantagem através da distinção entre estados, processos e eventos mentais, de um lado, e mente, de outro. Embora se tenha ciência de algumas das próprias percepções, estados, e processos mentais, a unificação deles numa única entidade é uma hipótese. A mente, definida como aquilo que *tem* esses estados, eventos e processos, é um postulado teórico, que, segundo Armstrong, o materialista identifica (de fato) com o cérebro organizado como o sistema nervoso central, com estados, processos e eventos cerebrais.

Essa distinção não é livre de questionamentos, e talvez se justifique pela forma como a identidade da mente foi expressa por esses teóricos, já que os objetos a serem identificados eram “estados, processos e eventos mentais” e a expressão utilizada para identificar o cérebro com a mente aparecia apenas como uma forma vaga e sem exatidão. Pode-se ver isso nas primeiras frases do artigo “The Identity Theory of Mind” (2000), no qual Smart afirma que a

teoria da identidade não precisa sustentar que a mente é idêntica ao cérebro, após apresentar a definição da teoria em termos de processos e estados mentais e cerebrais. Em termos materialistas, esta distinção é bem vinda porque elimina um termo geral, mente, em prol de termos que se relacionam a ocorrências mais factuais, concretas, e melhor definidas, processos, eventos e estados mentais. Assim, poderia-se dizer que não há mentes, apenas estados, eventos e processos mentais.

A observação dos próprios estados e processos mentais não é negada, mas há uma restrição ou distinção entre a sua observação e a atribuição categorial do que está sendo observado. Isso se manifesta no percebê-los *como* mentais. Denominar esses objetos como sendo mentais seria uma teoria ou quase-teoria, na opinião de Armstrong. Ele os chama de entidades semiobservacionais e semiteóricas. São, pois, duas coisas distintas, a coisa percebida ou percepção e como a percebemos.

Dizer do que se sente que é uma coisa ou outra é categorizar essa sensação. Assim, ter a sensação de um objeto pode ser descrito da seguinte forma: (i) “uma esfera esverdeada de superfície levemente rugosa sobre uma superfície plana de cor branca”; e também como (ii) “uma laranja *não-madura* sobre uma mesa branca”. Embora a descrição (i) já faça uso de categorias como formas geométricas, cores e texturas para classificar (nomear) os objetos observados, pode-se notar que a descrição (ii) vai além e apresenta categorias que estão em um nível no qual não há relação direta com os fenômenos observados, como no primeiro caso, mas já se aplicam categorias mais amplas. “Uma esfera esverdeada de superfície levemente rugosa” é vista como “uma laranja não-madura” e “uma superfície plana de cor branca” como “uma mesa branca”. Espero que este exemplo simples mostre a categorização à qual a sensação está submetida a cada momento em que fazemos descrições de sensações.

Esta distinção é reforçada na forma do argumento da mulher decapitada que foi criado por Armstrong e que Smart cita algumas vezes.

A mágica da mulher decapitada funciona do seguinte modo: um mágico utiliza como cenário um palco escurecido com cortinas pretas e uma mulher veste uma roupa que contrasta com esse fundo escuro. O truque de mágica consiste em fingir decapitar a mulher de alguma forma utilizando como elementos para simular a decapitação um capuz preto que cobre a cabeça da mulher, cujo corpo é bem iluminado. Sendo assim, a cabeça da mulher fica imperceptível. Com isso, *não vendo a cabeça da mulher* (não tendo a percepção de sua cabeça), o espectador é levado a *ver uma mulher sem cabeça*.

Esse argumento faz uso da descrição de um truque de mágica para mostrar o erro categórico que se comete ao considerar que a mente não é física, ao qual Armstrong e Smart

dão o nome de ilusão metafísica (Smart, 2006). Embora o argumento tenha sido apresentado como resposta a um problema mais específico, à percepção de que os estados mentais não são físicos, há uma semelhança pois nos dois casos os fenômenos são vistos *como*. A pessoa percebe seus processos internos *como* processos mentais e uma mulher da qual ela não vê a cabeça, *como* uma mulher sem cabeça.

Esse argumento pode ser considerado como um complemento à falácia fenomenológica apresentada por Place, que é definida como

O erro de supor que quando o sujeito descreve sua experiência, quando ele descreve como as coisas parecem, soam, cheiram, têm um sabor ou são sentidas por ele, ele está descrevendo as propriedades literais dos objetos e eventos em um tipo peculiar de cinema ou tela de televisão internos, usualmente referidos na literatura psicológica moderna como o “campo fenomenal”. (Place 1970 [1956], p. 49)

Seria, então, necessário distinguir a forma como as coisas aparecem, a parte fenomenológica, da real natureza das coisas. Vemos que essas duas falácias (erros metafísicos) são complementares entre si e podem ser usados em favor da identidade entre cérebro e mente.

6.1.2 Teoria Causal da Mente, de David Lewis

David Lewis defendeu uma teoria da identidade de tipo, desenvolveu uma teoria causal da mente que possibilitasse essa *identidade lógica* e que respondesse a algumas objeções. Na parte introdutória ao artigo “An Argument for the Identity Theory” (1983), que foi originalmente publicado em 1966, ele expressa seu argumento em termos de papel causal. Assim, o papel causal é a característica definitiva e analiticamente necessária de qualquer experiência, sendo esta um universal, e, como uma matéria de fato, esse mesmo papel causal pertence a certos estados físicos. Como consequência, esses estados físicos são identificados com essa experiência. Em outras palavras, o que vai definir uma experiência, ou estado mental, é o papel causal que ela desempenha. Desse modo, as coisas que exercem esse mesmo papel causal podem ser identificadas com essa experiência.

Seguindo as orientações de Place e Feigl, e as tentativas de Smart e Armstrong de definir a mente e o cérebro de modos lógicos independentes, Lewis desenvolve a prova lógica

ao estabelecer a identidade dos estados mentais e dos estados físicos com o mesmo *objeto*. Lembre-se que as formas como Smart e Armstrong definiam os estados mentais e cerebrais permitiam que as entidades objeto de identificação fossem simplificados, pois não haveria dois tipos diferentes de coisas para serem identificadas, já que as fórmulas de definição e relato eram neutras. A definição de Lewis é muito próxima da de Armstrong e ele parte, nesse artigo, da identificação dos estados mentais e cerebrais com o mesmo papel causal, como segue: (1) “Estado mental M = o ocupante do papel causal R (por definição de M)”. (2) “Estado neural N = o ocupante do papel causal R (pela teoria fisiológica)”. (3) “Portanto, estado mental M = estado neural N (por transitividade de $=$)”. Lembre-se que foi por definição que Armstrong identificou os itens mentais com estados causais, portanto, sua forma de identificação pode ser inserida na prova de Lewis e, portanto, segundo seu ponto de vista, também ser provada logicamente.

Acho proveitoso fazer uma breve exposição dos passos lógicos em que essa identidade é feita em (Lewis, 1980 [1972], pp. 208 e seguintes). Essa exposição será bem literal e pode ser considerada quase como uma grande citação, para que não se percam a legitimidade na construção lógica do argumento.

Em primeiro lugar, a psicologia popular é tomada como sendo uma teoria que, como todas as teorias, é composta de termos teóricos – termos-T – que são os termos postulados, introduzidos pela teoria, aqueles que ela define, e os outros termos – termos-O – que são os demais termos presentes na teoria, também chamados de termos velhos. Os termos-T são introduzidos por uma teoria e significam o que a teoria, que poderia ser entendida como uma história na qual os termos-T desempenham papéis específicos, afirma que eles significam.

Os termos teóricos dessa teoria podem ser representados por uma n -upla, um vetor contendo o termos teóricos da teoria. Lewis faz uso da noção de descrição definida para satisfazer esse vetor com apenas uma realização para cada um dos termos-T, ou seja, cada termo é realizado por uma e apenas uma entidade no mundo, se há mais de uma realização ou nenhuma realização para algum desses termos-T, então a teoria é falsa. A teoria também pode ser aproximadamente realizada ou então pode conter erros que podem ser corrigidos, quanto a isso não há objeção. Assim, os termos-T são *nomes* que satisfazem uma teoria e, quando são encontradas entidades que satisfazem unicamente a teoria, ela é verdadeira.

Agora fica fácil entender os passos do argumento.

Suponha uma nova teoria T , introduzida pelos termos $t_1 \dots t_n$, dado n um número natural qualquer, que são seus termos-T. Todos os outros termos na teoria são os termos-O. A teoria T é apresentada na forma de uma sentença longa, podendo ser uma conjunção. Essa

teoria fala de entidades que são nomeadas na teoria pelos termos-T que elas são ocupantes dos *papéis causais* apresentados pela teoria. A teoria T, então é postulada como

$$T[\mathbf{t}].$$

quer dizer, uma teoria que introduz os termos $\mathbf{t}$. Observe que a expressão de $\mathbf{t}$ em negrito significa que $\mathbf{t}$ é uma n -upla, ou seja, um vetor com n elementos.

Com a substituição dos termos-T por variáveis livres, $x_1 \dots x_n$, obtém-se a fórmula na qual apenas os termos-O aparecem,

$$T[\mathbf{x}].$$

Veja que, nesse passo, considerando-se a filosofia da mente, cada conceito mental seria substituído por uma variável. A teoria da primeira parte da teoria causal da mente, de Armstrong, estaria nesta posição.

O próximo passo é o estabelecimento empírico da identificação das variáveis com entidades que satisfazem o papel causal em questão. Cada n -upla de entidades que satisfazem a fórmula anterior é chamada de realização de T . Se há uma realização para T , então isso leva a uma quantificação existencial da fórmula anterior, denominada *sentença de Ramsey*. A sentença de Ramsey afirma simplesmente que existe pelo menos uma realização para T .

$$\text{Ex } T[\mathbf{x}].$$

Neste caso, no texto do artigo também se usa o E normal no lugar do quantificador de existência, que seria invertido. O próximo passo é identificar a teoria para satisfazer a descrição definida, ou seja, para que exista apenas uma realização da teoria:

$$\text{E}_1 \mathbf{x} \ T[\mathbf{x}].$$

O próximo passo é transformar a sentença de Ramsey modificada na sentença *condicional de Carnap*. Essa sentença vai afirmar que, “se T for realizada, então os termos-T nomeiam os componentes de alguma realização de T ”. Observe que este passo não utiliza a sentença anterior, a sentença modificada de Ramsey.

Se $\text{Ex } T[\mathbf{x}]$, então $T[\mathbf{t}]$.

A realização de uma teoria serve para instanciar os termos da teoria. Os termos- T são substituídos até que se encontre uma realização para a teoria, a partir desse momento passa-se a ser legítimo utilizá-los enquanto tal. Assim, a teoria da identidade não excluiria os conceitos mentais, o que torna-a uma teoria não redutível, nesse sentido, e não eliminativista. A exclusão dos conceitos mentais ocorre em um momento determinado do processo de identificação e não como uma postura geral ou ontológica.

O próximo movimento consiste em modificar a sentença de Carnap de modo a condicionar o teorema à existência de uma única realização da teoria, junto com uma outra fórmula para cobrir os casos de a teoria não ser realizável.

Se $\text{E}_1\mathbf{x } T[\mathbf{x}]$, então $T[\mathbf{t}]$,

Se $\sim \text{E}_1\mathbf{x } T[\mathbf{x}]$, então $\mathbf{t} = *$.³⁵

Esse par de sentenças é considerado equivalente à seguinte sentença, que afirma que $\mathbf{t}$ é a realização da fórmula com variáveis.

$\mathbf{t} = \text{o } \mathbf{x } T[\mathbf{x}]$.

Considero as implicações da teoria de Lewis mais fortes do que as decorrentes das teorias de Smart e Armstrong, porque Lewis compreende que essa teoria permite a substituição dos termos. Não ficou claro se isso equivale à tradução de sentido, se as sentenças com um dos termos teria o mesmo sentido que a sentença com os outros termos. Mas é certo que esta operação é feita com os termos de uma mesma teoria. Se for nesse sentido, então utilizar t ou x não vai alterar o sentido da teoria.

Vamos supor que todo o movimento que foi feito até este ponto serviu para demonstrar que o “estado mental $M = \text{o ocupante do papel causal } R \text{ (por definição de } M\text{).}$ ” Dando prosseguimento ao argumento, “vamos supor que uma outra n -upla, especificada de outra forma que a forma especificada pela entidades que satisfazem T , realmente realizam T .” Ou seja, há um outro conjunto de entidades que também satisfazem a teoria, mas essas entidades

³⁵ Segundo nota na versão desse artigo no livro (Lewis, 1999, p. 255), o $*$, que no artigo é representado por uma estrela de cinco pontas, significa que cada t_i do vetor $\mathbf{t}$ não tem denotação.

parecem ser diferentes das entidades anteriores, ou seja, são especificadas de outra forma, uma teoria cujos termos- T são do tipo $\mathbf{r}$.

$$T[\mathbf{r}].$$

Essa sentença com o termos- T $\mathbf{r}$ é posta independentemente de T , com $r_1 \dots r_n$ sendo ou termos- O ou termos de uma outra teoria. Veja que, se o mesmo processo que foi aplicado a $T[\mathbf{t}]$ for aplicado a $T[\mathbf{r}]$, embora Lewis não apresente esse desenvolvimento, tem-se que

$$\mathbf{r} = o \ x \ T[\mathbf{x}].$$

o que implica a identidade. Mas Lewis não desenvolve a teoria dessa forma, ele apresenta a aceitação de uma redução fraca para T , da qual se segue que:

$$\mathbf{t} = \mathbf{r}.$$

Uma outra forma de redução que também pode assumir a forma da crença em uma única realização de T , uma *redução forte*, tal que, se $\mathbf{r}$. É a afirmação de que qualquer que seja $\mathbf{x}$, se $\mathbf{x}$ realiza T , então $\mathbf{x} = \mathbf{r}$. Dessa forma, assume-se que T tem uma única redução, $\mathbf{r}$, e qualquer que seja a realização de T , essa realização será igual a $\mathbf{r}$. E daí se conclui que $\mathbf{r}$ é a realização de T .

$$(\mathbf{x}) (T[\mathbf{x}] \text{ sse } \mathbf{x} = \mathbf{r})^{36},$$

$$\mathbf{r} = o \ x \ T[\mathbf{x}].$$

Esta é a expressão lógica da identidade entre mente e cérebro, porém é uma identidade que tem uma única realização. Agora serão apontadas mais algumas características da teoria de Lewis.

No artigo “An Argument for the Identity Theory”, Lewis apresenta algumas respostas a objeções levantadas contra a teoria da identidade. A primeira tratou da objeção que atribui

³⁶ Neste caso, o quantificador universal é expresso como “(x)” na notação utilizada nesse artigo de Lewis e adotada nesta dissertação.

características dos objetos da experiência aos estados físicos, por exemplo, achar que a experiência da cor púrpura deve ser ela mesma púrpura. A segunda diz respeito à não localização (por necessidade analítica) da experiência, quer dizer, que por definição os estados mentais não são localizados no espaço, e localização dos eventos neurais, ambos considerados como particulares abstratos por Shaffer³⁷, ele chega a uma ideia interessante sobre a natureza dos eventos neurais. Segundo consta nessa passagem, eventos neurais também são abstratos.

O seu argumento consta de duas premissas, a primeira diz que experiências são definidas por papéis causais, e a segunda, diz respeito à adequação explicativa da física. No artigo, o filósofo trata de cada uma dessas premissas e da conclusão isoladamente.

A primeira parte do argumento determina o papel causal de uma experiência, qualquer que seja ela, como a sua característica definitiva. Segundo Lewis, esse papel causal *definitivo* “é expresso por um número finito de condições que especificam suas causas e efeitos típicos sob várias condições.” Ainda, segundo suas palavras, essa forma de definir as experiências é uma “elaboração” e “generalização” da teoria de Smart sobre os relatos de experiência através da fórmula “O que está acontecendo em mim é como o que acontece em mim quando ...” [*What is going on in me is like what is going on in me when ...*]. Para Lewis, a região pontilhada pode ser preenchida com compostos lógicos; ou com causas e efeitos típicos da experiência definida, podendo ser outras experiências; e não se aplica apenas a relatos de primeira pessoa. Essas formas de relato e suas relações de neutralidade serão analisadas no capítulo 7.

Nesta teoria, as “experiências”³⁸ são reais e interdefinidas, ou seja, na definição de uma experiência x , outras experiências $x_1, x_2, \dots, x_n$ ³⁹ são incluídas, de modo que essa relação é considerada crucial na definição de seus papéis causais. Essas experiências são organizadas em famílias que são mutuamente interdefinidas. Os membros de uma família são determinados pelo isomorfismo de seus papéis causais. Assim, se x ocupa um papel causal definitivo para uma experiência dentro da família X , isso se deve à pertinência de x ao isomorfismo causal dessa família de experiências. Uma família pode ser entendida como um conjunto de elementos que satisfazem um mesmo critério. Esse isomorfismo é expresso como o padrão de relações causais que a experiência possui com outras experiências internas à família, e relações causais com ocorrências externas à família, como estímulos e respostas.

37 Filósofo a quem Lewis responde e considera no artigo.

38 Na teoria de Lewis, experiência também é identificada com estado, mas neste parágrafo faço amplo uso da palavra experiência, pois é desta forma que Lewis se expressa, mas não haveria grandes problemas se fosse utilizada a palavra estado.

39 Segundo Lewis, não há impedimento de que o conjunto de condições que estão na definição de uma experiência seja infinito, por causa da recursão, mas ele opta por limitá-las, por achar que não há perigo de proceder dessa maneira. Neste caso, $x_1, x_2, \dots, x_n$ referem-se às experiências envolvidas na definição, e considerou-se n como sendo um número natural, portanto, uma sequência finita de experiências.

Com isso, é estabelecido um critério de identidade, todas as experiências dentro de uma família são identificadas com suas contrapartes no isomorfismo causal da família.

Apesar de usar esse critério de identidade, a definição de experiências, assim como provavelmente a determinação do papel causal que determina uma família, não é exaustiva com respeito a todas as ocorrências de uma experiência. Lewis identifica experiência como o estado numa relação causal típica, ou seja, aquele que é tipicamente causado de certas formas e que tipicamente dá origem a certos efeitos. Então, a teoria pressupõe que haja certas ocorrências que não são abarcadas pela definição de uma experiência, casos que são excepcionais.

Estes são pontos que podem ser considerados suficientes para entender a filosofia desenvolvida por Lewis no contexto estudado. Para finalizar, convém observar as distinções entre as causalidades defendidas por Armstrong e por Lewis. Salvo engano, Armstrong define a causalidade presente em sua teoria como sendo ordinária, eficiente, no estilo humeano, sendo esta a mesma causalidade que opera no mundo físico. Ele também afirma que “a causa e seu efeito são ‘existências distintas’, de modo que a existência da causa não implica logicamente a existência do efeito, ou *vice-versa*” (Armstrong, 1993, p. 83). Além disso, ele afirma que “se uma sequência é causal, então ela cai sob alguma lei”. Lewis define causalidade em termos contrafactuais. Se um dado evento *e* ocorre em todo mundo possível em que é precedido por um evento *c*, então *c* é causa de *e*. Neste caso, a ocorrência de *e* é dependente da existência de *c*. A abordagem de Lewis é muito mais complexa do que a forma simplificada apontada, mas importa saber que a distinção entre a noção de causalidade nos dois casos está não só na explicação em termos contrafactuais proporcionada por David Lewis, que tem fortes implicações ontológicas, como na dependência que os dois eventos envolvidos na cadeia causal possuem um com o outro, relação de dependência que foi rejeitada por Armstrong.

7 Neutralidade

No artigo “Sensations and Brain Processes” (1959), J. J. C. Smart, ao responder à *terceira objeção* à teoria da identidade, apresenta o que viria a ser um dos principais problemas da teoria da identidade entre os filósofos materialistas australianos: prover a teoria com um relato ou descrição dos estados mentais que seja neutra entre o dualismo e o materialismo (entre outras teorias, com exceção explícita do epifenomenalismo). A linguagem empregada na descrição do mental deveria ser neutra entre o materialismo e o dualismo, ou seja, não específica, e, a partir desse texto de Smart, tanto Armstrong quanto Lewis se empenharam posteriormente em desenvolver uma forma de definição dos estados mentais que mantivesse ou alcançasse essa neutralidade. Essa busca por uma definição tópico neutra faz com que essas teorias sejam denominadas também como sendo “da forma translacional”, pois há tradução dos conceitos mentais ordinários, tais como crenças, sensações e desejos, para uma linguagem tópico neutra.⁴⁰

A terceira objeção é apresentada da seguinte forma:

“Objeção 3. Mesmo se as Objeções 1 e 2 não provam que sensações são algo mais além de processos cerebrais, elas realmente provam que as qualidades das sensações são alguma coisa além de processos cerebrais. Isto é, pode ser possível escapar de afirmar a existência de processos psíquicos irreduzíveis, mas não é o caso de não ter mais de afirmar a existência de propriedades psíquicas irreduzíveis. Pois suponha que nós identifiquemos a Estrela da Manhã com a Estrela Vespertina. Então deve haver algumas propriedades que implicam logicamente aquela de ser a Estrela da Manhã, e propriedades muito distintas que acarretam aquela de ser a Estrela Vespertina. Novamente, deve haver algumas propriedades (por exemplo, aquela de ser uma clarão amarelo) que são logicamente distintas daquelas na história física.” (Smart, 1959, p. 148)

Essa objeção argumenta em favor da existência de propriedades não físicas, ou seja, que há propriedades das sensações que não são processos cerebrais, não são materiais, se formos um pouco mais além nesse argumento. Assim, mesmo que as sensações sejam identificadas a processos cerebrais, ainda haverá propriedades que são irreduzíveis a esses processos. Aqui, há uma oposição entre processos e propriedades. Nessa objeção, o dualista parece abrir mão da irreduzibilidade dos processos mentais em favor da irreduzibilidade das propriedades, mesmo quando a teoria visa identificar especificamente os processos físicos e processos mentais (sensações).

Haveria propriedades que dizem respeito às sensações e propriedades que diriam

40 Cf. Borst (1970, p.20).

respeito aos cérebro. No argumento dado por Smart, utilizou-se o caso das estrelas Vespertina e Matutina que, na realidade, são o mesmo corpo celeste visto no céu em horários diferentes do dia. Assim, é afirmado que há propriedades àquele corpo enquanto se atribui a ele o nome de Estrela Vespertina e que não fariam sentido ao serem atribuídas ao mesmo corpo quando é chamado de Estrela Matutina. Assim, haveria propriedades mentais que não conservariam seus sentidos se fossem atribuídas a estados e processos cerebrais.

Pode ser um pouco inútil cogitar o motivo de Smart não ter respondido essa objeção dizendo que, a partir do momento que se descobre a identidade dos dois corpos, ou seja, que ambas as estrelas são uma e a mesma estrela, o planeta Vênus, então as propriedades da Estrela Vespertina são as propriedades do corpo em uma determinada posição no espaço em um horário especificado, enquanto as propriedades da Estrela Matutina são as propriedades do corpo em uma outra posição e horário especificados. Ou seja, que quando se descobre a identidade, se atribui ambas as propriedades a um único corpo. A insatisfação presente nesse tipo de resposta é que os nomes do corpo estão em jogo e é enquanto propriedade de nomes que o problema se apresenta. Desse forma, mesmo aceitando que a Estrela Matutina é o mesmo corpo que a Estrela Vespertina, não faz sentido atribuir ao nome Estrela Matutina às propriedades pertencentes ao nome Estrela Vespertina.

Ser tópico neutro nesse sentido é não especificar a natureza dos estados mentais e ela está intimamente ligada ao fato de uma pessoa não sentir seus próprios estados e processos mentais como materiais. Assim, ele dirá que os relatos dos processos mentais não especificam as propriedades dos processos relatados. Nos relatos sobre sensações não há comprometimento com propriedades físicas ou mentais, eles são *tópico neutros* nesse aspecto.

De acordo com Smart, o termo “tópico neutro” foi usado por Gilbert Ryle para se referir a palavras lógicas ou quase-lógicas, que quando ouvidas seletivamente numa conversa não discriminam o assunto que está sendo abordado, se se fala sobre geologia, matemática, assuntos de fazenda ou quaisquer outros (SMART, 2006, 2000).

Essa neutralidade assume algumas formas, em primeiro lugar será analisada a neutralidade da teoria causal.

Na teoria causal, o que vai determinar um estado e processo mental é a relação causal que ele mantém com estímulos, comportamentos e outros processos e estados mentais. Sendo determinado por essa relação sem qualquer referência a estados e processos físicos, a teoria causal mantém-se neutra, podendo-se atribuir um dado estado mental a um espírito incorpóreo, um corpo humano ou um termostato.

Segundo Armstrong, na mágica metafísica da mulher sem cabeça, descrita

anteriormente na seção 6.1.1, a base para essa ilusão seria a aquisição de crenças negativas.

Fazendo uma substituição dos termos, com respeito aos próprios estados e processos mentais, a não percepção deles como sendo materiais gera a falsa conclusão de que eles não são materiais. Nas palavras de Armstrong, “falha de ser ciente de materialidade será naturalmente interpretada como ciência de imaterialidade.”

A importância da neutralidade para a teoria da identidade surge também na seguinte afirmação em Smart (2006, p. 160), de que “a idéia de tópico neutro é absolutamente vital para a teoria da identidade”; e também em Armstrong (1993, p. 77-8):

que explicação verbal ou ‘definição ostensiva’ pode ser dada do significado desta palavra [‘mente’] sem implicar uma partida de um ponto de vista fisicalista do mundo? Este parece ser o grande problema, ou, de alguma razão, um grande problema, enfrentado por uma teoria de estado central.

A neutralidade, no caso, entre teorias da mente como o dualismo e o materialismo, torna-se, então, um critério para o que esses filósofos consideram uma boa abordagem da mentalidade, como é o caso de Armstrong. Segundo ele, para uma teoria da mente ser satisfatória teria que dar conta de quatro itens⁴¹; (i) permitir a possibilidade lógica da mente desencorporada; (ii) incapacidade lógica de existência independente dos estados mentais; (iii) o problema da intencionalidade; (iv) abarcar, além dos eventos mentais internos, comportamento e disposição para se comportar.⁴² Esses itens são oriundos de várias teorias da mente distintas e uma boa teoria, pois, teria que os preservá-los.

Também Smart expressa a consistência da linguagem tópico neutra com diferentes teorias sobre o problema mente corpo, como as materialistas e mesmo as imaterialistas. Desde “Sensations and brain processes”, quando responde à terceira objeção à teoria da identidade, ele utiliza o argumento da linguagem tópico neutra. Essa objeção foi assumidamente a mais difícil de responder e para a qual ele não ficou bem certo quanto ao sucesso da resposta, tanto que a fórmula foi retomada e sofreu modificações algumas vezes, especialmente no artigo “Reports of Immediate Experiences” (Smart, 1971), quando ela vai sendo avaliada e reconstruída.

Na resposta à terceira objeção, Smart também adentra um assunto que lhe é constante, a filosofia das cores. Apresenta a ideia de percipiente humano normal e, com isso, a definição de cor como poder para evocar certos tipos de respostas discriminatórias. O conceito de

41 (Armstrong, 1993, p. 76 e 91-92)

42 Itens já abordados no capítulo 5.

percepiente humano normal é utilizado para introduzir o conceito de normal e desenvolver a fórmula tópico neutra, que dará início à busca da fórmula tópico neutra por Armstrong e Lewis, porque é nesse tema que ele põe em prática o que chama de teoria da identidade, a neutralidade de sua linguagem.

Algumas vezes essa neutralidade empresta à teoria da identidade um caráter de teoria fraca de identidade mente-cérebro. Tomemos como exemplo dessa posição o termo “cérebro”, que Smart considera num sentido amplo designando não apenas o cérebro humano ou animal, como também aqueles dos habitantes de Alfa Centauro, com todas as diferenças físico-químicas e anatômicas, ou o de um robô do futuro, com *microchips* em vez de massa cinzenta (Smart, 1989, p. 153). Essa variabilidade parece depender de uma definição funcional do cérebro, posto que a semelhança buscada não parece ser física, mas caracterizada pela função que o cérebro exerce, possibilitando, na teoria da identidade, um aspecto do funcionalismo que costuma ser fortemente utilizado contra a identidade entre mente e cérebro, a múltipla realização do mental. Quanto a isso, ele afirma, em Smart (1991, p.169), que o teórico da identidade também poderia concordar com os vários meios físicos de realização dos estados mentais definidos funcionalmente, mas que ao teórico da identidade cabe observar algumas semelhanças relativamente abstratas entre esses tipos físicos e que, esse teórico, continuaria defendendo uma identificação tipo-tipo ao menos com relação aos estados mentais de uma mesma pessoa. Observe-se que ele defenderia essa espécie de identidade com um grau não especificado de abstração, pois fala de “semelhanças relativamente abstratas”, mas não se sabe qual é o seu limite, onde um grau de abstração deixa de ser adequado à teoria.

Outras linguagens tópico neutras que foram desenvolvidas na filosofia da mente, como no funcionalismo de Putnam, com a linguagem de máquina de Turing, e na descrição de Lewis que é baseada na lógica de Ramsey e que foi vista na seção 6.1.2.

Vale salientar uma outra função da linguagem tópico neutra, que diz respeito a lógica de uma identidade contingente. Pois, para identificar estados e processos cerebrais e mentais contingentemente, faz-se necessário não incorrer em circularidade. Por isso, especificamente, Armstrong inserirá esta forma de definição da mente em sua teoria (Armstrong, 1993).

8 Homem máquina e múltipla realização

No livro *Philosophy and scientific realism* (Smart, 1963), no capítulo intitulado “Man as a Physical Mechanism”, Smart desenvolve de forma mais aprofundada a hipótese de homem máquina que ele já havia apresentado no artigo “Sensations and Brain Processes” (1959). Ela é apresentada na forma da tese: o homem é um mecanismo físico; e também é expressada como “o homem é uma máquina”. O presente capítulo apresenta um corte histórico da concepção de máquina, a mecanicista cartesiana, a máquina computacional de Turing, a máquina físico-química de Smart. Ele mostra um pouco da evolução mecanicista, e a evolução do próprio conceito de máquinas, os avanços tecnológicos e teóricos que possibilitaram o desenvolvimento da inteligência artificial e o levantamento de questões sobre a própria condição do ser humano enquanto tal. Por que isso? Porque a neutralidade desenvolvida nas fórmulas da teoria da identidade não são impedimento para a implementação da mente em formas não orgânicas de vida, por exemplo.

Considerar o homem como uma máquina não é algo novo em filosofia, Descartes fez isso, e La Mettrie escreveu o livro *L'Homme Machine*. Nesse aspecto, Descartes é interessante. Na “Quinta Parte” do “Discurso do Método”, ele faz um resumo do seu tratado de Física, que o filósofo adiou publicar por receio de ser desaprovado pelas autoridades, conforme relato na “Sexta Parte”. Os motivos de tais receios eram o fato de uma opinião de física de um autor, cujo nome e obra não são revelados, ter sido desaprovada, na qual Descartes afirma não ter encontrado nada que pudesse prejudicar a religião ou o Estado ou que o impedisse de escrevê-las; e receava que algumas de suas próprias ideias também viessem a ser censuradas.

8.1 Autômatos cartesianos

Descartes é considerado um dualista de substância. Não é tese deste trabalho defender Descartes como um materialista, principalmente um que defenda a teoria da identidade, mas em algumas passagens ele parece atribuir mais ao cérebro do que talvez um dualista atribuiria.⁴³ Além disso, há a identidade que ele estabelece entre as regras da Mecânica e as da

43 Quanto a isso, podemos ler também: (i) Borst (1970, p. 14): “Mesmo Descartes estava preparado para

natureza. Primeiro, ainda na “Sexta Parte” do *Discurso do Método*, ele afirma que “o espírito depende tanto do temperamento e da disposição dos órgãos do corpo, se é possível encontrar algum meio que torne os homens mais avisados e mais hábeis do que foram até aqui, creio que é na Medicina que se deve procurá-lo.” (Descartes, 1979 [1637], p. 61) Essa crença na medicina expressada em todo o *Discurso do Método* não é muito diferente da que os filósofos materialistas contemporâneos, especialmente os teóricos da identidade e eliminativistas, sustentam com relação às Neurociências. Com isso, são quase quatrocentos anos de esperança no amadurecimento da medicina.

Será conveniente apresentar as duas passagens, a que trata do corpo humano e a que trata da criação de máquinas de homens e animais.

Explicara assaz particularmente todas essas coisas no tratado que pretendi outrora publicar. E, em seguida, mostrara nele qual deve ser a estrutura dos nervos e dos músculos do corpo humano, para fazer que os espíritos animais, estando dentro, tenham a força de mover seus membros: assim como se vê que as cabeças, pouco depois de decepadas, se remexem ainda, e mordem a terra, não obstante não mais sejam animadas; quais mudanças se devem efetuar no cérebro, para causar a vigília, o sono e os sonhos; como a luz, os sons, os odores, os sabores, o calor e todas as outras qualidades dos objetos exteriores nele podem imprimir diversas idéias por intermédio dos sentidos; como a fome, a sede e as outras paixões interiores também lhe podem enviar as suas; o que deve ser nele tomado pelo senso comum, onde essas ideias são acolhidas; pela memória, que as conserva, e pela fantasia, que as pode modificar diversamente e compor com elas outras novas, e pelo mesmo meio, distribuindo os espíritos animais nos músculos, movimentar os membros desse corpo de tão diversas maneiras, quer a propósito dos objetos que se apresentam a seus sentidos, quer das paixões interiores que estão nele, que os ossos se possam mover, sem que a vontade os conduza. (Descartes, 1979 [1637], p.59-60)

Na outra passagem, continuando na mesma página dessa edição, Descartes também trata de máquinas. Enquanto uma máquina que reproduzisse um animal sem razão, ou seja, com seus órgãos e sua figura, não poderia ser distinguida do próprio animal, uma outra máquina que imitasse da mesma forma um ser humano, “imitando tanto nossa ação quanto moralmente possível”, seria reconhecida por dois meios. O primeiro, seria a utilização de palavras ou “outros sinais” em composições para comunicação de pensamentos. Este parece um uso autônomo e criativo da linguagem, não simplesmente uma programação de palavras apropriadas para determinadas situações. O segundo meio diz respeito à multiplicidade de

admitir, com consistência duvidosa, um papel especial, em relação com operações mentais, ao cérebro e especialmente à glândula pineal”. (ii) Feigl (1970 [1960], p.35): “O quebra-cabeça crucial e central do problema mente-cérebro, pelo menos desde Descartes, consistiu no desafio de dar uma explicação adequada da relação das ‘sensações brutas’ [raw feels], tanto quanto dos fatos mentais (intenções, pensamentos, volições, desejos, etc.) ao processo neurofisiológico correspondente.” De qualquer forma, não estou defendendo que Descartes seja materialista.

formas significativas que o ser humano age de acordo com as situações. Assim, a razão nos permite agir de modos diversos e apropriados em diversas circunstâncias, enquanto realizar tal máquina com disposições corporais para agir nas variadas ocorrências da vida seria “moralmente impossível”.

Com respeito ao conceito de máquina em Descartes e da relação do corpo e da alma, na qual o cérebro exerce um papel chave, pode-se tomar o livro *As Paixões da Alma* como boa fonte de referência. Nesse texto, a palavra máquina reaparece constantemente para designar o corpo. Além disso, os animais seriam apenas máquinas, mesmo aqueles que reproduzem palavras como os papagaios.

8.2 Computadores e máquina de Turing

Desde então, as máquinas mudaram devido a avanços tecnológicos e conceituais, como os desenvolvimentos Física, Química, da cibernética e da ciência da computação, etc. e o que Descartes considerava algo “moralmente impossível” de se conseguir foi apresentado novamente, por Alan Turing no artigo “Computing Machinery and Intelligence”, como critério para julgar uma máquina inteligente. Turing propõe uma nova formulação para a questão “máquinas podem pensar?”. Turing descreve uma nova forma para o problema, derivada de uma brincadeira de festa chamada “jogo da imitação”. O jogo tem três integrantes, um homem (A), uma mulher (B) e um interrogador (C), que pode ser de qualquer sexo e que fica num quarto separado dos outros dois. Ele os conhece por X e Y, e no final do jogo terá que dizer uma das duas afirmações, ou que “X é A e Y é B” ou que “X é B e Y é A”.

Como meio ideal de comunicação ele sugere uma tela para mensagem instantânea (*teleprint*), ou também as respostas poderiam ser escritas, ou tipografadas, ou mesmo respondidas através de um intermediário, pois esse tipo de meio evitaria identificação por meio da tonalidade das vozes. O papel de um dos jogadores, B, é ajudar o interrogador dizendo a verdade, enquanto cabe ao jogador A, atrapalhá-lo, ou seja, passando-se por B.

Desse jogo, Turing retira as novas questões que substituirão a anterior “máquina pode pensar?”, que são: “O que acontecerá quando uma máquina tomar o lugar de A no jogo?” e “O interrogador decidirá erroneamente tão frequentemente quanto quando o jogo é jogado por um homem e uma mulher” (Heil, 2004, p.112). Heil nomeia esse procedimento de definição operacional, pois assume a forma de um teste para saber se o conceito se aplica.

Segundo Turing, as vantagens dessa nova formulação é que ela delimita precisamente as capacidades intelectuais das capacidades físicas de um homem. Ele fala “capacidades”, mas talvez a palavra mais adequada seja “características”, pois ele justifica isso com respeito à possibilidade de se criar uma pele ou carne artificiais que sejam muito parecidas com as humanas e ao fato de o interrogador ouvir, tocar ou ver os outros competidores.

Note-se que, no novo jogo da imitação, é a máquina que imita o homem e que assume o papel de atrapalhar o interrogador. Um motivo para isso é que o inverso seria muito evidente, isto é, seria muito fácil de ser distinguido, pois um humano faz cálculos de aritmética mais lenta e menos acuradamente, por exemplo.

Turing propôs que a máquina para jogar esse jogo seria o computador digital, ou eletrônico, e que em vez de se deter sobre os computadores de sua época, seria melhor imaginar um computador que se daria bem nesse jogo. Esse computador digital faria tudo (todos os cálculos) o que um computador humano é capaz de fazer, mas este sofre algumas restrições para a analogia ser bem sucedida. A primeira diz que se supõe que o computador humano segue regras fixas, das quais “não tem nenhuma autoridade de desviar em qualquer detalhe”. A segunda, que há um livro com essas regras, que é alterado cada vez que ele muda de tarefa. E a terceira, que há um suplemento ilimitado de papel para se fazerem os cálculos.

O modelo de computador digital é composto de três partes: armazenamento (que reproduz o papel no qual o computador humano faz seus cálculos, ou mesmo na memória, quando ele faz cálculos de cabeça; parte da área de armazenamento é reservada para o livro de instrução que o computador humano segue, e é chamada de tabela de instruções); unidade executiva (a parte que executa as várias operações envolvidas no cálculo); e um controle (esta parte verifica se as instruções do livro de instrução, que ele chama de tabela de instrução, estão sendo seguidas corretamente e na ordem certa).

No computador digital, a tabela de instruções será a descrição do programa, que nada mais é que os passos seguidos para executar uma tarefa.

O computador digital de Turing é uma máquina de estado discreto determinista, ou seja, a tabela de instruções é a descrição de estados discretos que podem ser diferenciados entre si e enumerados por números inteiros, e das relações desses estados para cada momento em que ela recebe um dado determinado. Sendo determinista, a partir do conhecimento dos estados da máquina em um certo momento, é esperado que, para determinadas informações que ela recebe, *inputs*, é sempre possível determinar quais respostas ela dará, *output*, seguindo as instruções da tabela.

A máquina retornará uma resposta enquanto o número de passos para a sua realização

for finito. Nos casos que ela demorar para retornar a resposta, o motivo para tal comportamento pode ser tanto a necessidade de um longo tempo para a execução da computação, quanto o caráter infinito da resposta. Pode-se ilustrar esses casos com dois exemplos, primeiro, multiplicações envolvendo números com n casas decimais, sendo n um número natural muito grande: a máquina levará um tempo grande para retornar o resultado, que poderá ser retornado dado o caráter finito dessa operação. Um segundo exemplo pode ser ilustrado de modo bem simples, como o maior número natural n ou o maior número primo x . Dado ser sempre possível encontrar um número natural que seja igual a $n + 1$, à princípio, a máquina operará por tempo infinito. Para que esses exemplos sejam autênticos, faz-se necessário explicitar que o tipo de resposta dado seria da forma “sim” ou “não”.

Uma outra característica importante dessa máquina diz respeito ao fato de ela poder imitar qualquer outra máquina determinística, sendo, portanto, equivalente, caracterizando, assim, sua universalidade. Assim, um computador digital poderia ser programado para realizar operações distintas apenas com a implementação de uma tabela de instruções, sem ser necessário construir um novo computador para cada atividade a ser realizada. De certa forma, os computadores atuais operam de acordo com uma máquina universal de Turing, enquanto servem para implementar vários programas diferentes, inclusive aquele que serviria para imitar o homem no jogo da imitação.

Em sua comparação com os seres humanos e outras máquinas, a sua capacidade de armazenamento (memória) é um aspecto que a diferencia e beneficia, pois os equipamentos que serviram de modelos na indução científica do que se considera uma máquina tinham uma memória limitada e restrita, enquanto os computadores de sua época que estavam operando em Manchester teriam uma capacidade de $2^{165.000}$ estados devido à capacidade de armazenamento que, por sinal era bastante limitada, dado que era correspondente a uma folha de papel comum (Turing, 1950. p. 441).

Para concluir este tópico, será apresentado o exemplo de máquina de Turing, presente no artigo, de uma roda que estala girando 120° por segundo, podendo ser parada por uma alavanca que pode ser operada de fora, acendendo uma lâmpada em função de sua posição. A posição da roda é descrita por um dos seguintes estados q_1 , q_2 ou q_3 . Há os sinais de *input* (entrada) i_0 e i_1 , que descrevem a posição da alavanca. E, da combinação do último estado da máquina e do *input*, tem-se a determinação da tabela de instruções que é dada por:

		Último Estado		
		q_1	q_2	q_3
Input	i_0	q_2	q_3	q_1
	i_1	q_1	q_2	q_3

Os *outputs* são descritos pela tabela:

Estado	q_1	q_2	q_3
Output	o_0	o_0	o_1

Uma descrição para o funcionamento desse sistema seria a seguinte. A posição da alavanca i_0 faz a máquina mudar do estado atual para o estado contíguo, lembrando que os estados representam posições da roda na qual o estado q_1 está entre os estados q_2 e q_3 .

Uma simulação desse funcionamento seria este:

1. Se o último estado da máquina for q_1 , abaixar a alavanca corresponde ao *input* i_0 e, fará com que a máquina passe para o estado q_2 , gerando como resposta o *output* o_0 , ou seja, a lâmpada não acende;
2. Após a realização de 1., a máquina agora está no estado q_2 , sendo este o último estado da máquina. Quando ela recebe o *input* i_1 , ou seja, quando a alavanca é levantada, a máquina continua no estado q_2 , sendo que seu *output*, representa um estado desligado da lâmpada.
3. Se a alavanca é baixada, após 2., estando a máquina no estado q_2 , ela irá para o estado q_3 , produzindo o *output* o_1 , para o qual a lâmpada se acende. Levantar a alavanca, i_1 , neste estado, fará com que a lâmpada continue acesa, pois a máquina continuará no mesmo estado.

8.3 Aprendizagem de Máquina

No artigo “Computing Machinery and Intelligence” (1950), Turing responde à objeção de Lady Lovelace segundo a qual as máquinas farão apenas aquilo que são programadas para fazer, programação aqui tendo o sentido de fazer o que se manda. Na formação da mente de um adulto, Turing considera que há três fatores: (a) o estado inicial da mente, no nascimento; (b) a educação à qual a pessoa foi submetida; e (c) outras experiências às quais a pessoa foi submetida mas que não são consideradas como educação (ele não especifica quais seriam estas experiências). Tendo feito esta distinção, em vez de se programar ou simular a mente de um adulto, o que requeria, segundo seus cálculos, uma capacidade de armazenamento de cerca de 10^9 dígitos binários, *bits*, para se jogar satisfatoriamente o jogo da imitação, ele sugere que se produza um programa que simule a mente de uma criança.

A sua descrição da mente de uma criança é a de um caderno novo, no qual há poucos mecanismos e muitas folhas em branco que seriam preenchidas com as experiências apropriadas para se formar um adulto. Turing também esclarece que o que ele chama nesse modelo de “mecanismo” seria quase sinônimo de escrita; em suas palavras: “Mecanismo e escrita são, do nosso ponto de vista, quase sinônimos” (p. 456). Além disso, essa pequena quantidade de mecanismos seria um facilitador de sua programação. Observe-se que os objetos de comparação com os programas e suas estruturas são o cérebro infantil ou o cérebro adulto, tanto em relação à capacidade de armazenamento quanto à sua capacidade de *operação*.

Assim, sua nova máquina seguiria o seguinte esquema:

Estrutura da máquina criança	=	Material hereditário
Mudanças	=	Mutações
Julgamento de quem faz o experimento	=	Seleção natural

Um motivo para se considerar essa versão dada por Turing para uma máquina que simule a mente humana é sua interpretação biológica, a qual podemos observar no esquema acima e na avaliação desses processos como biológicos.

Assim, ele compara a evolução desse programa com a evolução biológica natural⁴⁴,

⁴⁴ Turing não faz distinção entre evolução e evolução natural, neste caso eu acrescentei essa palavra para

afirmando que a sobrevivência do mais apto é um método lento de se medir vantagens, e a aplicação desse mecanismo na programação pelo experimentador poderia ser agilizada. Uma forma de se agilizar o processo evolutivo é por meio do controle da mutação aleatória, ou seja, a pessoa que faz o experimento pode rastrear a causa de uma dada fraqueza e pensar em um tipo de mutação que a melhore.

Um outro processo de ensino segundo o qual uma máquina criança pode ser construída, e que foi também apresentado por Turing, é a punição e a recompensa. Isto consiste em tornar menos repetitivos os eventos que precedem os sinais de punição e aumentar a probabilidade de repetição daqueles que são recompensados. Infelizmente, Turing não dá exemplo da implementação desses algoritmos, dizendo apenas que fez alguns experimentos com tal máquina com algum sucesso, mas que seu método não foi muito ortodoxo. Apesar disso, é possível presumir que essa máquina-criança é quase que exclusivamente linguística, pois há passagem na qual ele apresenta as suas limitações em comparação com as ações de uma criança envolvendo, principalmente, partes do corpo como pés, olhos, braços. Por exemplo, ele utiliza uma ilustração na qual não se poderia ensinar a maquininha a encher um balde de carvão. Embora não desconsidere totalmente a possibilidade de avanços na engenharia que possibilitassem a criação de uma máquina inteligente que representasse uma criança, ele expressa sua preocupação com o que poderia-se chamar de “*bullying*”, dizendo que não se poderia mandar a criatura para a escola sem esperar que as outras crianças gozassem muito dela. É impressionante ele apresentar essa preocupação, porque é pressuposto que a máquina não possui emoção, tanto que os seus canais de comunicação são não-emocionais. Por isso, poderia ser irrelevante esse comportamento das crianças em relação às máquinas. Posteriormente, ele analisa alguns exemplos de implementação de aprendizado e imperativos e novamente seus elementos apresentam um caráter lógico e linguístico, enfatizando o fato de sua máquina não possuir corpo e de que por essa razão os imperativos têm caráter intelectual. Em outras palavras, sua base de conhecimentos seria proposicional e armazenada na forma de linguagem simbólica. A estrutura ou mecanismo pode variar de um programador para outro, podendo haver uma estrutura lógica fixa ou adquirida por meio da aprendizagem.

No caso de haver um sistema lógico prefixado, não poderia ser descartado como sendo a sua forma de implementação, pois ele desenvolve este método um pouco mais largamente de modo que se harmoniza com suas ideias de aprendizado. Assim, utilizando uma máquina

diferenciar a parte que ele compara a evolução no seu programa à evolução no sentido que biólogos e o público em geral usa o termo.

na qual o sistema lógico não precisa ser aprendido, boa parte do espaço de armazenamento seria ocupado por definições e proposições. Estas poderiam ser classificadas em vários tipos, como esses expressos por Turing: fatos bem estabelecidos, conjecturas, teoremas matematicamente provados, afirmações de autoridade, e expressões tendo a forma lógica de proposição mas que não têm valor de crença (proposições imperativas).

Turing apresenta o funcionamento dos imperativos da seguinte forma:

- i. algumas proposições são descritas como ‘imperativas’;
- ii. proposições imperativas classificadas como ‘bem-estabelecidas’ levariam automaticamente à ação;

Assim, por exemplo, supondo que um professor diga à máquina ‘Faça sua lição de casa agora’, isso pode causar a inclusão da proposição “o professor diz ‘Faça sua lição agora’” entre os fatos bem estabelecidos. Supondo que haja um outro fato bem estabelecido na memória que diz “Tudo o que o professor diz é verdade”, combinando-se as duas proposições tem-se que ‘Faça sua lição de casa agora’ é incluída em fatos bem estabelecidos e, devido à forma de construção da máquina, isso significa que essa atividade já foi iniciada. Pode-se incluir regras para operação lógica, como as sequências segundo as quais as operações devem ser realizadas, ou indicações de escolher as melhores ações, como aquelas mais rápidas, por exemplo. E, nesse processo, elas poderiam, inclusive, fazer inferências científicas.

Uma consequência de uma máquina-criança bem treinada seria a imprevisibilidade de seu comportamento, mesmo que não tenha sido eliminada a possibilidade de sua arquitetura determinística. Da forma como Turing apresenta a máquina criança, a objeção de Lady Lovelace torna-se estranha, porque ele aponta para a possibilidade de uma máquina fazer novas inferências que são imprevisíveis. Com isso, ele afirma que a maioria dos programas que são implementados numa máquina produzirão como resultado comportamentos para os quais não havia expectativa e que poderiam ser considerados completamente aleatórios.

A aleatoriedade não é vista apenas como uma emergência nessa máquina, mas Turing chega a apresentá-la como um método, posto que, para adivinhar um número em um dado intervalo, pode ser menos oneroso partir da escolha aleatória, tentando encontrar um número mais próximo daquele que se deseja descobrir, a seguir a sequência numérica. E novamente ele relaciona esse método com a biologia, segundo a qual os processos evolutivos funcionam de modo análogo.

Essa relação entre inteligência artificial, biologia e teoria da identidade também pode

ser visualizada com a analogia da “pele de uma cebola”, expressa na página 454. Dada a beleza dessa passagem, vale a pena fazer uma citação de todo o trecho nas palavras do autor:

Considerando-se as funções da mente ou do cérebro, nós encontramos certas operações que podemos explicar em termos puramente mecânicos. Essa dizemos que não corresponde à mente real: é um gênero de pele que devemos despir se vamos alcançar a verdadeira mente. Mas então, no que resta o que nós encontramos é uma outra pele para ser despida, e assim por diante. Procedendo dessa forma, sempre alcançamos a mente ‘real’, ou eventualmente alcançamos a pele que não contém nada? No último caso, toda a mente é mecânica.

Turing não apenas criou o computador virtual, como também aproximou a inteligência artificial da cognição humana, especialmente relacionada ao desenvolvimento cognitivo, embora essa parte do seu artigo não receba a mesma atenção que a sua máquina virtual. A relação dos filósofos considerados nesta dissertação com a inteligência artificial aparece de uma forma ou de outra, geralmente por referência explícitas a outras formas de “máquinas”, como os robôs, e também com a própria extensão da palavra máquina para nomear o corpo humano.

Pode-se fazer uma distinção dentro da teoria da identidade entre os filósofos que estenderam seus conceitos aos robôs (objetos da cibernética e da Inteligência Artificial), e aqueles que se opuseram a essa extensão. Um exemplo dos filósofos da teoria da identidade que se posicionou contra é Herbert Feigl. Na seção V-E do artigo “The ‘Mental’ and the ‘Physical’”, Feigl se posiciona contra a exibição por parte de robôs e andróides de todas ou da maioria das características do comportamento humano. A razão para esse impedimento seria a constituição física dessas máquinas, pois a hipótese de essas máquinas serem constituídas das mesmas proteínas que formam os sistemas nervosos, isso poderia fazer com que elas pudessem apresentar essas características comportamentais. Porém, Feigl não considera plausível essa hipótese de haver a mesma constituição física, e nem a presença de uma linguagem intersubjetiva dentro de uma máquina de silício, por exemplo.

8.4 Homem como um mecanismo físico-químico

A importância da noção de mecanismo físico-químico na teoria da identidade deve-se ao fato de ser ela o que possibilita à teoria da identidade a múltipla realização do mental, além

de compatibilizar essa teoria da mente com as outras áreas científicas como biologia, química e física. Em outras palavras, esse ponto de vista com relação ao ser humano poderia não apenas estar de acordo com a ciência, mas também ser visto como um produto científico. As seguintes palavras de Smart extraídas do artigo “Sensations and Brain Processes” sustentariam essa ideia: “Parece-me que a ciência está cada vez mais dando-nos um ponto de vista segundo o qual os organismos são capazes de serem vistos como mecanismos físico-químicos.”

A noção de homem máquina não é irrelevante na distinção entre funcionalismo e teoria da identidade, porque esta teoria se apoia nessa redução fisicalista, mas o mesmo não ocorre com a primeira. Putnam, por exemplo, nega expressamente que seres humanos sejam máquinas (Putnam, 1975 [1960], p. 364). Assim, enquanto a teoria da identidade apresentada por Smart nivela todos os seres ao mesmo nível físico-químico e assim dá suporte teórico à múltipla realização do mental, um outro funcionalismo, como o de Putnam, lidará com a diferença, os seres humanos não seriam máquinas, mas o mental seria multiplamente realizado.

No livro *Philosophy and scientific realism*, Smart apresenta a tese de que o homem é um mecanismo físico. Uma das críticas à teoria da identidade é a capacidade de múltipla realização que provém das teorias de máquina funcionalistas, porém essa tese é apontada por Smart como um exemplo em prol da teoria da identidade, já que se uma máquina, robô, é capaz de pensar, então um ser humano como uma máquina também o será. Além disso, ele não é contra a ideia da inteligência artificial forte. No livro *Nosso lugar no universo*, ele trata da questão dos robôs e considera que é possível que uma tal máquina seja consciente, inclusive, uma dessas possibilidades é a similaridade entre o cérebro de um robô e um cérebro humano. Em Smart (2004) também há uma passagem na qual ele afirma

As pessoas frequentemente pensam que se um robô pudesse escanear seus próprios processos isso não poderia significar que o robô era consciente. Isso é um apelo à mera ‘intuição’ da qual deveríamos ser suspeitos. Poderíamos reverter o argumento e dizer que porque consciência é ciência de ciência, o robô é consciente. Acho que essa resposta realmente vai contra minha própria intuição (o que quer que isso seja) mas eu deveria ser tão suspeito de minhas próprias intuições como eu sou das dos outros. (Smart, 2004, p. 48)

Smart aparece bem receptivo sobre as previsões de desenvolvimento da inteligência artificial, inclusive na sua abordagem do problema de Gödel presente no artigo “Gödel theorem, Church’s theorem, and mechanism”. Aqui ele afirma que, com a aprendizagem de

máquina, um computador poderá resolver problemas de engenhosidade na hora de provar e resolver problemas matemáticos, em razão do desenvolvimento no campo de reconhecimento de padrões na área de inteligência artificial. Podemos citar mais algumas passagens nas quais ele apresenta esse otimismo com relação ao desenvolvimento dessa área:

Esse conhecimento ‘intuitivo’ é realmente indutivo, e poderia ser adquirido, tenho sugerido, por uma máquina. [...] A despeito disso, tenho esperança de que meu próprio argumento, que faz uso de uma máquina indutiva, possa, talvez, ser de interesse pelo motivo que ela é provavelmente mais proximamente relacionada com o modo no qual o cérebro humano trabalha. (Smart, 1961, pp. 109 – 110)

Para buscar uma solução para o problema de Gödel e construir a ideia abstrata de máquina intuitiva, Smart utiliza a teoria de máquina de Turing, que foi tratada acima. Apesar disso, em Smart (1959b), ele apresenta uma visão um pouco mais negativa, afirmando que um robô pode fazer discriminações de cores, porém as discriminações de cores ou falta delas são “características altamente idiossincráticas da espécie humana [...] da fisiologia humana” (Smart, 1959b, p. 40).

No livro *Philosophy and scientific realism*, pode-se dizer, Smart faz um esboço de uma teoria neurocomputacional⁴⁵. Ao tratar do problema da criatividade e da liberdade humana, Smart argumenta em favor de uma visão determinista do ser humano que não contradiz esses aspectos. Ele argumenta que “a razão, a persuasão e a ameaça podem ser fatores causais no comportamento humano” (Smart, 1961, p. 123) e aconselha o leitor a imaginar uma máquina complexa que pudesse ser influenciada por esses fatores, informando que tal máquina deveria ser determinística. Então ele faz uma relação com a forma como algumas pessoas usam a indeterminação quântica como argumento em favor do livre arbítrio. A objeção levantada contra este argumento tem como fundamento a redundância no cérebro humano, onde vários neurônios são responsáveis pelo comportamento. Citando Smart:

[...] Essa espécie de redundância é de fato, em uma extensão bem menor, construída dentro de certas máquinas computacionais, quando é necessário que elas devam

45 O primeiro modelo matemático de um neurônio foi desenvolvido em 1943 pelo neurofisiologista, filósofo e poeta americano McCulloch e pelo lógico Walter Pitts, portanto já havia essa teoria quando Turing e Smart escreveram seus textos. Esse modelo é bem simples, o neurônio pode estar em dois estados ativado ou desativado. O neurônio possui um determinado número de entradas, cada uma com um peso atribuído. Para que haja ativação, ele soma todos os produtos dos pesos e das entradas que recebe e, se o resultado dessa soma for maior que um valor de ativação estipulado, o neurônio é ativado. (Bittencourt, 2006)
Infelizmente, neste trabalho não há lugar para escrever mais sobre este tema, embora possa-se defender uma relação da teoria identidade da forma como é definida por Smart e a neurocomputação ou neurociências da forma como ela é desenvolvida hoje. De certa forma, essas áreas podem servir de argumento em favor da teoria da identidade desenvolvida pelos materialistas australianos.

trabalhar em um modo confiável mesmo que tenham componentes não confiáveis. Se nós pensamos no cérebro humano como uma máquina, é plausível pensar nele como uma máquina determinística. (Smart, 1961, 123)

Essa passagem evidencia e problematiza a noção de homem máquina adotada por Smart, que se aproxima de uma visão na qual a múltipla realização pode ser derivada, não se apresenta mais como um contraexemplo para a teoria da identidade da forma como ele a sustenta. A aproximação entre homem e máquina, nos aspectos morais, se evidencia na seguinte passagem: “Se uma máquina é capaz de fazer escolhas e de ser influenciada por punição e recompensa, então ela tem em essência tudo que é preciso para o livre arbítrio” (Smart, 1961, 124).

Em uma outra passagem no mesmo capítulo, Smart refuta um argumento segundo o qual as máquinas computacionais substituem os homens em apenas algumas atividades que este executa, portanto o homem é diferente de computadores com respeito a muitos outros aspectos. Smart, então, afirma que esse argumento não prova “que um artefato não poderia ser construído para realizar todas as funções, tanto computacionais quanto não-computacionais, que os homens realizam” (Smart, 1961, p. 109). Esse argumento é complementado com a afirmação de que, mesmo se uma máquina computacional não for capaz de desempenhar todas as funções humanas, mesmo assim, a consideração do homem como uma máquina física não é invalidada.

Essas passagens mostram uma abertura para outras formas de mentalidade, também presente em Smart (1995). Elas são interessantes e geram algumas reflexões sobre o caráter da teoria da identidade e também sobre o compromisso de J. J. C. Smart com essa teoria, pois pode-se questionar sua preocupação e sua adesão a essa teoria. De qualquer forma, este livro apresentou um conteúdo bastante amplo, no qual a teoria da identidade (a teoria de processo cerebral) estava presente. A mudança que Smart afirma que a teoria da identidade sofreu ao longo dos anos foi a extensão da identificação para todos os estados, processos e eventos mentais feita por Armstrong, não sendo, então um desvio da teoria original, mas uma ampliação da mesma. Além disso, a teoria da identidade mantém um diálogo de acordo com as ciências, especialmente as físicas (trata-se de uma teoria reducionista na qual a química, a biologia e a psicologia são redutíveis à física).

Quando Armstrong trata da teoria da identidade em (1984, p. 160), ele apresenta o materialista, e neste caso ele está falando daquele que identifica a mente ao cérebro numa identificação tipo-tipo, como alguém para quem a *noção* de inteligência artificial não é uma

dificuldade. Essa “noção de inteligência artificial” compatível com tal materialismo pode ser expressa como (i) “não é o caso que ‘as únicas mentes fisicamente possíveis são de natureza neurofisiológica’”; (ii) “dado ‘um programa de computador suficientemente complexo e sofisticado, o computador controlado por esse programa poderá exibir inteligência’, num sentido nem metafórico nem secundário”; e (iii) “computadores com inteligência levam a crer na possibilidade da construção de mecanismos com alguns dos seus estados internos podendo ser identificados com estados mentais”. Esse argumento de Armstrong serve para complementar as relações entre teoria da identidade e inteligência artificial e mostrar o quanto esses filósofos aceitaram uma tal visão da inteligência e da consciência para formas não orgânicas de *vida*.

9 Relações entre a teoria da identidade e o funcionalismo

A relação entre a teoria da identidade e a teoria causal da mente não é de oposição, pelo menos a princípio. No desenvolvimento desta última, tanto David Armstrong quanto David Lewis tinham a intenção de defender a teoria da identidade de objeções e críticas e torná-la mais aceitável e inteligível. Em outras palavras, as teorias de Armstrong e Lewis foram inicialmente propostas como argumentos em favor da verdade da teoria da identidade, para comprovar e desenvolvê-la. Além disso, sempre houve uma interação e influência entre Smart, Armstrong e Lewis. Smart admite com bastante frequência a influência das ideias de Armstrong em sua filosofia, e este desenvolveu algumas ideias que eram, em certa medida, uma continuação das obras de Smart. Além disso, a presente dissertação mostra a continuação teórica que há entre essas teorias, que o desenvolvimento do funcionalismo nesses termos foi uma implicação dos aspectos da própria teoria da identidade, suas características e exigências.

Por exemplo, Lewis (1983 [1966], p. 99), afirma que aqueles, ele inclusive, que aceitam a hipótese materialista, na qual fenômenos físicos têm apenas explicações físicas, *devem* aceitar a teoria da identidade. E esse “devem” está explícita e expressamente posto em oposição a escolher livremente. Para ele, não é uma questão de liberdade, elegância ou economia aceitar essa teoria, como afirma Smart, mas um dever de todo materialista. Além disso, o filósofo trata especificamente da identidade de tipo, distinguindo entre universais e particulares abstratos. Na nota número 1 desse artigo, ele explicita essa distinção da seguinte forma: há dois tipos de universais com os quais está comprometido, o primeiro é universal no sentido mais amplo e diz respeito a uma experiência que “acontece a muitas pessoas em muitos tempos”, e a segunda, que é mais restrita, pode ser expressa como a experiência de “algum tipo definido que pode pelo menos ser comum a diferentes pessoas em tempos diferentes”.

Lendo os dois autores, é possível ver que a referência por eles adotada para a teoria da identidade é o livro *Philosophy and Scientific Realism*, de J. J. C. Smart, por isso os argumentos utilizados na teoria causal defendem uma identidade que se diferencia do que se poderia esperar de uma tal teoria, ou seja, a teoria assume certos graus de similaridade em vez de identidade quando trata de identificar tipos de estados e processos mentais.⁴⁶ Pode-se explicitar o que se poderia esperar de uma teoria da identidade assumindo-se a concepção

⁴⁶ A identidade entre processos mentais e cerebrais é ontológica, as semelhanças se aplicam às experiências perceptivas, ao que os outros, à fenomenologia da percepção. São estes que são semelhantes, como aparece na fórmula tópico neutra de Smart.

clássica, segundo a qual uma identidade entre estados e processos mentais e cerebrais será sempre uma identidade necessária. Nos textos de Smart, especialmente no livro *Philosophy and Scientific Realism*, o conceito de identidade utilizado no relato da percepção fenomenológica não só se aproxima do de similaridade, como as percepções são definidas em função dessa similaridade, *qualia* como pontos em um espaço de similaridade. Por esse motivo, a teoria causal desenvolvida por esses dois filósofos aparenta defender, justificar ou reforçar mais o que foi expresso no livro *Philosophy and Scientific Realism* do que a teoria da identidade à qual os críticos se opõem. Para evitar essa percepção, é preciso ter em mente que todos os estados, inclusive as experiências, são identificadas com estados cerebrais, mas percebemos, experienciamos em termos de similaridade.

A especificação de identidade geralmente utilizada pela crítica para exemplificar essa teoria, “dor é igual à ativação de fibras C”, na qual eles fazem o leitor subentender que a dor seria identificada a uma fibra específica e que se há alguma alteração na fibra que foi usada na identificação, por exemplo, ou ela se altera ou não há tal fibra num sistema no qual se está fazendo a identificação, então a identidade deixa de ser válida. Mas, como foi mostrado, tal conclusão não se segue, já que a teoria da identidade é especificada de forma bastante abstrata e a identidade entre esses processos é contingente, havendo espaço para a múltipla realização.

Um outro aspecto que se pode salientar é a ênfase dada por Armstrong à distinção entre a sua e a teoria da identidade, ou pelo menos, entre a parte dois e a parte três do seu livro *A Materialist Theory of the Mind*. Assim, no último capítulo, ele “re-enfatiza” a independência lógica entre as duas partes, ou seja, entre a análise lógica e a identidade que se segue, afirmando que nada que ele disse na parte dois (a análise conceitual dos conceitos mentais feita nessa parte do livro) acarreta a verdade de uma teoria materialista da mente. E segue afirmando que “O melhor que pode ser dito da Parte Dois é que nela o caminho para a identificação foi suavizado.”⁴⁷ A falta de uma relação lógica entre as duas teorias, que ele afirma nessa parte do livro, com a ressalva de haver observações incidentais feitas por ele em algum momento no texto vinculando-as, torna o núcleo central de sua teoria independente da teoria da identidade. Essa relação lógica diz respeito à primeira/segunda partes do livro, a análise lógica, implicar logicamente a teoria da identidade. O núcleo lógico da teoria de estado central bem como da própria fórmula de relato de experiência de Smart não implicam logicamente a verdade do materialismo, nem consequentemente da teoria da identidade, segundo Armstrong e Smart, embora Lewis tenha tentado provar o contrário.

Finalmente, a teoria da identidade não está em confronto sequer com o funcionalismo

47 Citação feita anteriormente [fazer um ponteiro para a citação anterior]

de máquina, até certo ponto. Smart faz uso da fórmula tópico neutra, o que a aproxima muito das fórmulas funcionalistas, podendo ser tão abstratas quanto estas. Até mesmo a possibilidade de múltiplas realizações e as relações com a inteligência artificial estão presentes na teoria de Smart. Pode-se perceber que a influência de Putnam e da cibernética aparecem muito claramente no livro *Philosophy and scientific realism*, quando ele desenvolve a ideia de homem máquina que já estava presente no artigo “Sensations and brain processes”.

O artigo Smart (1959b), cuja publicação antecedeu o artigo de Smart mencionado, foi escrito em oposição às ideias de um artigo de Putnam publicado em 1956 na *Philosophical Review*. Segundo Smart, nesse artigo, Putnam afirmava que noções como ‘exatamente a mesma cor como’ são, por definição, “relações transitivas, reflexivas e simétricas que implicam mas não são implicadas pela relação ‘indistigível de forma de’ e que, por isso, é necessária (Smart, 1959b, p. 39). Smart argumenta em favor dessa noção como sendo contingente, baseado no fato de as palavras para cores serem definidas ostensivamente.

A busca por uma linguagem neutra já estava presente na teoria da identidade, em função inclusive da necessidade de desenvolver uma identidade contingente entre mente e cérebro. Com isso, o desenvolvimento do funcionalismo causal a partir da teoria da identidade, da modo como ela foi formulada por esses filósofos, aparece como um processo natural, sem que haja rupturas muito rigorosas ou evidentes dentro da teoria da identidade. Por isso, certamente, o livro de Armstrong foi reclassificado mas mesmo assim ainda poder ser considerado uma teoria da identidade.

10 Conclusão

A exposição dos filósofos materialistas australianos feita neste trabalho mostra que as teorias da mente funcionalistas e teoria da identidade são teoricamente próximas. Essa proximidade não se deve apenas à reclassificação dos filósofos que defendiam uma versão da teoria da identidade, David Armstrong e David Lewis: a metodologia utilizada no desenvolvimento da teoria da identidade culminou na teoria causal da mente, que foi reconhecida a posteriori como uma teoria funcionalista.

Há duas formas de abordar o problema mente e corpo, a primeira diz respeito ao relato dos próprios estados mentais; a segunda diz respeito à definição desses estados. Como e o que se relata ao dizer “eu *tenho* uma dor”? Essa relação de possuir algo aparece nos textos de Place, Feigl, Smart, que são os primeiros, e exprime o realismo defendido por esses filósofos. A partir da tarefa de tratar esses processos como reais, Smart se preocupou em determinar a forma de relato de modo a permitir o desenvolvimento da teoria da identidade, uma forma/fórmula que tornasse a identificação possível. Houve, então, um empenho lógico/linguístico da parte desses filósofos, pois o que pode ser observado nos textos em questão, de Place a Lewis, é a presença e desenvolvimento de uma forma neutra e contingente de definição ou relato, forma esta que já estava presente em Place (1969 [1956]).

O artigo “Computing machine and intelligence”, de Alan Turing, do qual há uma breve exposição no capítulo 8, foi publicado em 1950. Embora não haja uma documentação *explícita* da influência de suas ideias nas obras de Smart, em 1959 ele escreveu um artigo no qual fazia uso nomeadamente de uma máquina de Turing para construir a teoria de uma máquina indutiva para lidar com os teoremas de Gödel e Church. Além de fazer uso da máquina de Turing, um outro assunto, a aprendizagem de máquina, à qual Turing já se referiu em 1950, e que também foi abordado no capítulo mencionado, a aprendizagem de máquina, foi utilizada no artigo de Smart em favor da máquina indutiva. Em Smart (1963), além de o filósofo fazer uma breve descrição da máquina de Turing, ele indica em nota para leitura um outro artigo desse cientista, “On computable numbers, with and application to the entscheidungsproblem”, e um pouco da teoria funcionalista de Putnam na qual os estados de máquina são definidos como estados lógicos e uma teoria da mente como estados lógicos.

Smart aceitou alguns aspectos da teoria proposta por Putnam, como pode ser visto no seu livro *Philosophy and Scientific Realism*. Por exemplo, em uma nota da página 2 desse

livro, ele afirma: “pensando na filosofia como reconstrução racional da linguagem, fui muito influenciado por Hilary Putnam e W. V. Quine.” Mas também se opôs a algumas de suas ideias, como pode-se ver em Smart (1959b, 1982). Incluiu em sua teoria e abordou os assuntos da cibernética, sem que houvesse conflitos aparentes para os postulados da teoria da identidade que defendia. Ao contrário, como aparece em alguns textos e foi citado nesse trabalho, Smart defendia uma concepção de homem máquina ou do sistema nervoso humano como uma máquina que não se sustenta ao fazer uma objeção à máquina de Turing. O conceito de máquina sofreu mudanças desde Leibniz, Descartes, La Mettrie. O conceito defendido por Smart se aproxima mais do conceito de Turing e, no período em que escrevia esses artigos, parecia está relacionado com esse modelo. Apesar disso, não se pode concluir que seu conceito de máquina fosse puramente turingiano em razão do seu fisicalismo, no qual o homem é considerado um mecanismo físico-químico muito complicado. Deve ser observado que, para Smart, a possibilidade de múltipla realização não aparecia como um problema para a teoria da identidade, mas uma confirmação de seu ponto de vista, de seu mecanicismo ou fisicalismo.

Smart, então, leu textos de Putnam e correspondeu de alguma forma à teoria que ele desenvolvia. Após escrever “Sensations and brain processes”, eles escreveu “Gödel’s Theorem, Church’s Theorem, and mechanism”, no qual faz uso da máquina de Turing e do conceito de máquina determinística como modelo para sua concepção de homem máquina, e em seguida publicou o livro *Philosophy and scientific realism*, texto que apresenta influência dessas teorias. Tanto Armstrong quanto Lewis mencionam especificamente esse livro em seus textos, o que mostra a influência que sofreram das ideias contidas nele.

Apesar disso, um outro aspecto que foi mencionado acima e que desempenhou um papel importante foi a busca de uma forma tópico neutra de definir ou relatar um estado mental. A necessidade dessa neutralidade pode ser atribuída a dois aspectos. O primeiro diz respeito à necessidade de se relatar os próprios estados mentais sem que se relacione a esse relato uma definição ontológica do que está sendo relatado. E o segundo, refere-se à possibilidade de identificação contingente. A primeira é uma forma de fuga da ilusão metafísica, quando não se sente o estado mental como sendo físico e se supõe que ele seja não físico.

O primeiro aspecto surge em Place (1969 [1956], p. 30) quando ele apresenta uma definição para uma pós-imagem nos seguintes termos:

Quando descrevemos a pós imagem como verde, não estamos dizendo que há

alguma coisa, a pós imagem, que é verde; estamos dizendo que estamos tendo a espécie de experiência que normalmente temos quando, e que aprendemos a descrever como, olhar um remendo verde de luz. (Place, 1969 [1956], p. 30)

Em Smart, essa forma de definição assume a fórmula: “o que está acontecendo em mim é como o que acontece quando...”, introduzida já no artigo “Sensations and brain processes” em resposta à objeção 3. Essa fórmula sofreu algumas alterações ao longo do tempo, como pode ser observado no capítulo 7 desta dissertação, mas foi um elo que relacionou os materialistas australianos. Armstrong irá mencionar esta fórmula, e criticá-la, ao criar (aperfeiçoar) sua própria fórmula.

Quando Armstrong cria sua fórmula, ele está sob influência não apenas da fórmula desenvolvida por Smart, mas também de uma necessidade de definir os conceitos mentais de forma que permitam uma identidade contingente entre os processos e estados mentais e cerebrais. A identidade contingente exige que os dois conceitos a serem identificados sejam independentes entre si. Armstrong então identifica o estado mental ao “estado da pessoa apto para causar uma certa espécie de comportamento” e alguns estados também como “estado da pessoa apto a ser causado por um estímulo”. As fórmulas alcançadas por Armstrong estão muito próximas das fórmulas dos funcionalistas e, além disso, podem ser atribuídas às necessidades da teoria da identidade: identificação contingente, ou seja, independência dos conceitos a serem identificados; e neutralidade, que está relacionada com o aspecto anterior, mas que também pode ser relacionada com a necessidade de neutralidade dos relatos psicológicos entre o materialismo e o dualismo.

Embora David Lewis tenha influenciado/aconselhado Smart na reformulação de sua fórmula tópico-neutra ao longo do tempo, ele deu prosseguimento à busca de uma definição neutra que possibilitasse a identificação contingente entre estados e processos mentais. Porém, contrariando os filósofos anteriores, ele tentou mostrar que a identificação entre a mente e o cérebro é uma questão lógica e não apenas científica. Com isso, ele apresentou uma demonstração lógica para essa identidade, como foi mostrado anteriormente. Além disso, a princípio, sua identidade é muito mais forte pois se baseia em uma descrição definida das teorias psicológicas. Claro que ele não afirma que a identidade mente-cérebro é realmente verdadeira, mas que o argumento para tal identificação é válido.

Pode-se afirmar sem o perigo de incorrer em erro grave, que a fórmula tópico neutra determinada por David Armstrong é simplificada por David Lewis. Na fórmula de Armstrong, o estado mental era definido como um estado que causa e que também é causado ou pode ser

causado, ou seja, um estado que desempenha um certo papel causal. Pode-se, então, dizer que Lewis tomou a fórmula de Armstrong simplificada e definiu o estado mental como “Estado mental M = o ocupante do papel causal R (por definição de M)”. Da mesma forma, o estado neural é definido como “Estado neural N = o ocupante do papel causal R (pela teoria fisiológica). Feito isso, Lewis procede à identificação “estado mental M = estado neural N (por transição de =)” (Lewis, 1999, pp. 248-249).

A dissertação se concentrou nesse movimento e nessas relações entre a teoria da identidade e o funcionalismo. No início da pesquisa havia apenas parte do conteúdo do capítulo 2, que tratava das confusões entre a teoria da identidade e o funcionalismo, ou seja, a reclassificação da sua própria teoria feita por Armstrong no livro *A materialist theory of the mind* e a indicação no artigo de Block.

Embora as teorias de Armstrong e Lewis sejam teorias causais, a questão da causalidade não foi aprofundada. As metafísicas desenvolvidas por esses dois filósofos é um pouco complexa e, embora eles se definam como fisicalistas, especialmente Lewis, os comprometimentos ônticos e ontológicos não são muito fáceis de se entender. Fazer um exame mais profundo da causalidade e de sua implicação no fisicalismo desses filósofos não seria muito aconselhável e nem necessário nesta fase da pesquisa.

REFERÊNCIAS

ARMSTRONG, D. M. The causal theory of the mind. In LYCAN, William G. (org.). **Mind and cognition: an anthology**. Malden: Blackwell, 1990.

_____. **A materialist theory of the mind**. 2ª ed. London: Routledge, 1993.

_____. The nature of mind. In BLOCK (org.) (1980), pp. 191-199.

_____. Reply to Smart. **Australasian Journal of Philosophy**. V. 84, Nº 2, pp. 177-178, junho 2006.

ARMSTRONG, D. M.; MALCOLM, Norman. **Consciousness & causality**. Oxford: Blackwell, 1984.

BITTENCOURT, Guilherme. **Inteligência artificial: ferramentas e teorias**. 3ª ed. rev. Florianópolis: Ed. da UFSC, 2006.

BLOCK, Ned (org.). **Readings in philosophy of psychology**. V. 1. Cambridge: Harvard University Press, 1980.

_____. What is functionalism? In BLOCK (org.) (1980), pp. 171-184.

BORST, C. V. (org.). **The mind-brain identity theory**. London: Macmillan; New York, St. Martin's Press, 1970.

CHURCHLAND, Paul. **Matéria e consciência: uma introdução contemporânea à filosofia da mente**. Tradução: Maria Clara Cescato. São Paulo: Editora UNESP, 2004.

DESCARTES, René. **Discurso do método; meditações; objeções e respostas; as paixões da alma; cartas**. (Os Pensadores) 2ª ed. Tradução: J. Guinsburg e Bento Prado Júnior. São Paulo: Abril Cultural, 1979.

EINSTEIN, Albert. **Essays in science**. New York: Philosophical Library, 1934.

FEIGL, Herbert. The 'mental' and the 'physical'. In FEIGL, Herbert; Maxwell, Grover,

SCRIVEN, Michael (org.). **Minnesota Studies in the Philosophy of Science, II**. Minneapolis: University of Minnesota Press, 1958, pp. 370-497.

_____. Mind-body, *not* a pseudo-problem. In BORST (org.) (1970), pp. 33-41.

FELDMAN, Fred. Identity, necessity, and events. In BLOCK (org.) (1980), pp. 148-155.

HEIL, John. **From an ontological point of view**. Oxford: Clarendon Press/Oxford University Press, 2003.

_____. **Philosophy of mind: a contemporary introduction**. 2ª ed. New York: Routledge, 2004.

HEMPEL, Carl G. The Logical analysis of psychology. In BLOCK (org.) (1980), pp. 14-23. [O texto foi publicado originalmente em 1935. O autor fez algumas modificações na versão inglesa, mas afirma que não houve mudança de conteúdo.]

HINTIKKA, Jaakko. On the different identities of identity: a historical and critical essay” in FLØISTAD, G. (org.) **Philosophical problems today**. V. 2. Dordrecht/Boston/London: Kluwer, 2005.

HUME, David. Investigação acerca do entendimento humano. In: BERKELEY, George; HUME, David. **Berkeley – Hume**. (Os Pensadores). Tradução de Anoar Aiex, João Paulo Gomes Monteiro, Armando Mora d’Oliveira. São Paulo: Nova Cultural, 1989.

KRIPKE, Saul A. **Naming and necessity**. Cambridge: Harvard University Press, 1980. [Primeira edição 1972.]

LEWIS, David. Psychophysical and theoretical identifications. In BLOCK (org.) (1980), pp. 207-215. [In LEWIS (1999), pp. 248-261. Esta edição contém uma notação melhor e notas mais esclarecedoras.]

_____. An argument for the identity theory. In LEWIS, David. **Philosophical Papers**. V. 1. New York: Oxford University Press, 1983. [O artigo foi publicado anteriormente em **Journal of Philosophy** 63 (1966): 17-25. Notas foram acrescentadas na edição de David M. Rosenthal, **Materialism and the Mind-Body Problem**. Englewood Cliffs: Prentice-Hall, 1971.]

_____. **Papers in metaphysics and epistemology**. Cambridge: Cambridge

University Press, 1999.

MORAVIA, Sergio. **The enigma of the mind: the mind-body problem in contemporary thought**. Tradução: Scott Station. Cambridge: Cambridge University Press, 1995. [Edição original: MORAVIA, Sergio. **Enigma delle mente**. Roma: Gius. Laterza & Figli Spa, 1986]

PLACE, U. T. Is Consciousness a brain process?. In *British Journal of Psychology*. V. 47:1, pp. 44-50, 1956. [In: BEAKLEY, Brian; LUDLOW, Peter (ed.) *The philosophy of mind: classical problems/contemporary issues*. 3a ed. Cambridge, Massachusetts; London: MIT Press, 1994, pp. 33-40. In: O'CONNOR, John (org.). *Modern materialism: readings on mind-body identity*. New York: Harcourt, Brace & World, 1969, pp. 21-31. In: BORST (1970) (org.), pp. 42-51. Na última página da primeira edição do artigo, de 1956, consta uma nota informando que o manuscrito foi recebido no dia 30 de novembro de 1954. Essa informação não é comentada nos escritos desses filósofos.]

_____. Materialism as a scientific hypothesis. In BORST (org.) (1970), pp. 83-86.

PRIEST, Graham. **An introduction to non-classical logic: from it to is**. 2ª ed. Cambridge: Cambridge University Press, 2008.

PUTNAM, Hilary. Minds and machines. In **Mind, language and reality: philosophical papers**, V. 2. Cambridge: Cambridge University Press, 1975. [Este artigo foi inicialmente publicado em 1960, no livro HOOK, Sidney (org.) **Dimensions of mind**. Nova York: Collier.]

RYLE, Gilbert. **The concept of mind**. London: Hutchinson's University Library, 1949.

SAVIN, Harris. Introduction: behaviorism. In BLOCK (org.) (1980), pp. 11-13.

SHIELDS, C. The first functionalist. In SMITH, J.C. (org.). **Historical foundations of cognitive science**. Dordrecht: Kluwer, 1990. pp. 19-33.

SMART, J.J.C. Incompatible colors. **Philosophical Studies**. V. 10, pp. 20-22, 1959b.

_____. Can biology be an exact science?. **Synthese**. V. 11, pp. 359-368, 1959a.

_____. Sensations and brain processes. **Philosophical Review**. V. 68, pp.141-156, 1959c. [In: O'CONNOR, John (org.) **Modern materialism: readings on mind-body identity**. New York: Harcourt, Brace & World, 1969, pp. 32-47. In: BORST (org.) (1970), pp. 52-66.]

_____. Gödel's theorem, Church's theorem and mechanism. **Synthese**. V. 13, pp. 105-110, 1961.

_____. **Philosophy and scientific realism**. London: Routledge & Kegan Paul, 1963.

_____. Materialism. In BORST (org.) (1970), pp. 159-170. [Texto original de 1963.]

_____. Further remarks on sensations and brain processes. In BORST (org.) (1970), pp. 93-94.

_____. Brain processes and incorrigibility. In BORST (org.) (1970), pp. 107-109.

_____. Reports of immediate experience. **Synthese**. V. 22, pp. 346-359, 1971.

_____. On some criticisms of a physicalist theory of colors. In CHENG, Chung-yin (ed.) **Philosophical aspects of the mind-body problem**. Honolulu: University Press of Hawaii, 1975. pp. 55-63.

_____. The content of physicalism. **Philosophical Quarterly**. V. 28, pp. 339-341, 1978.

_____. Metaphysical realism. **Analysis**. V. 42, pp. 1-3, 1982.

_____. Methodology and ontology. In GAVROGLU, K.; GOUDAROULIS, Y.; NICOLACOPOULOU, P. (eds). **Imre Lakatos and theories of scientific change**. Dordrecht: Kluwer, 1989, pp. 47-57.

_____. **Nosso lugar no universo: uma questão de espaço-tempo**. São Paulo: Edições Siciliano, 1991.

_____. 'Looks Red' and Dangerous Talk. **Philosophy**, v. 70, p. 545-554, 1995a.

_____. Postscript (1995). In O'CONNOR, Timothy; ROBB, David. **Philosophy of mind: contemporary readings**. pp. London: Routledge, 2003. pp. 133-137. [Segundo nota

dessa edição, este artigo foi originalmente intitulado “Postscript”. In MOSER, Paul K.; TROUT, J. D. (org.). **Contemporary materialism: a reader**. London: Routledge, 1995. pp. 104-106. Na bibliografia de Smart consta apenas esta edição do “Postscript”]

_____. Consciousness and awareness. **Journal of Consciousness Studies**. V. 11, n. 2, 2004, p. 41-50, 2004.

_____. The identity theory of mind. In **Stanford Encyclopedia of Philosophy**. Primeira publicação Jan 12, 2000; revisão substantiva em 18 de Maio de 2007. Disponível em: <<http://plato.stanford.edu/entries/mind-identity/>> Acesso em: 28 de ago. 2008, às 17h28.

_____. Comments on Hodgson. **Journal of Consciousness Studies**. V. 12, pp. 58-64, 2005.

_____. Ockhamist comments on Strawson. **Journal of Consciousness Studies**. V. 13, n. 10-11, pp. 158-162, 2006.

_____. Metaphysical illusions. **Australasian Journal of Philosophy**. V. 84, n. 2, pp. 167-175, jun. 2006.

STRAWSON, Galen. **Mental reality**. Cambridge: MIT Press, 1994.

TOLMAN, Edward. Cognitive Maps in Rats and Men. In **Psychological Review** 55 (1948): pp. 189-208.

TURING, Alan. Computing machine and intelligence. In **Mind**, New Series, V. 59, nº 236, out. 1950, pp. 433-60.

WETZEL, Linda. Types and tokens. In **Stanford Encyclopedia of Philosophy**. Primeira publicação em 28/04/2006. Disponível em <<http://plato.stanford.edu/entries/types-tokens/>> Último acesso em 26/10/2010.

WETZEL, Linda. **Types and tokens**: on abstract objects. Cambridge: MIT Press, 2009.