U.PORTO
FLUP FACULDADE DE LETRAS
UNIVERSIDADE DO PORTO

Mestrado em Filosofia
Filosofia Contemporânea

Realismo modal e conteúdo mental em David Lewis

João António Faria e Silva

2019

João António Faria e Silva

Realismo modal e conteúdo mental em David Lewis

Dissertação realizada no âmbito do Mestrado em Filosofia, orientada pelo Professor Doutor João Alberto Cardoso Gomes Pinto

Faculdade de Letras da Universidade do Porto

Setembro de 2019

Realismo modal e conteúdo mental em David Lewis

João António Faria e Silva

Dissertação realizada no âmbito do Mestrado em Filosofia, orientada pelo Professor Doutor João Alberto Cardoso Gomes Pinto

Membros do Júri

Professora Doutora Sofia Gabriela Assis de Morais Miguens Travis
Faculdade de Letras - Universidade do Porto

Professor Doutor João Alberto Cardoso Gomes Pinto
Faculdade de Letras - Universidade do Porto

Professor Doutor Mattia Riccardi
Faculdade de Letras - Universidade do Porto

Classificação obtida: 20 valores

Sumário

Declaração de honra

Declaro que a presente dissertação é de minha autoria e não foi utilizada previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Porto, 30 de setembro de 2019

João Silva

Agradecimentos

Tenho de agradecer, antes de mais, ao meu orientador, o Professor João Alberto Pinto, pelo empenho e disponibilidade no acompanhamento deste trabalho, e pelo incentivo que me deu durante todo este tempo. Além disso, foi quem despertou o meu interesse por alguns dos temas tratados nesta dissertação, através das estimulantes aulas de Filosofia da Mente e dos comentários que fez ao projeto de dissertação e à própria dissertação. É seguro dizer que sem essas aulas e comentários esta dissertação, a existir, teria tido um tema completamente diferente.

Agradeço também à minha família, especialmente aos meus pais, à minha irmã e aos meus avós, o incansável apoio e a importante compreensão. É de referir que foram os meus pais que me garantiram os meios monetários para eu realizar todo este percurso e, por isso, sem a ajuda deles, esta dissertação não existiria.

Finalmente, quero agradecer a todos os meus amigos e colegas que me ajudaram de uma maneira ou de outra, alguns pelos comentários e conversas sobre os temas tratados, outros pelo incentivo e companheirismo. Não posso esquecer, especialmente, o facto de a Diana Couto ter lido e comentado integralmente a versão provisória de dois capítulos desta dissertação.

Devo-vos a todos um sincero obrigado.

Resumo

O meu propósito nesta dissertação é apresentar as várias teses de David Lewis relativamente à ontologia da mente e ao tratamento adequado do conteúdo mental. Entre essas teses estão a sua variedade de identidade entre a mente e o corpo, a sua conceção de interpretação radical, a sua defesa de uma classe de atitudes intencionais irredutivelmente egocêntricas (ou em primeira pessoa) e a relevância do conteúdo restrito para a sua teoria geral da mente. Para completar esta tarefa, terei também de apresentar algumas das ideias metafísicas de D. Lewis que estão na base de algumas das suas propostas em filosofia da mente – nomeadamente, o realismo modal genuíno e a classificação de propriedades pelo seu grau de naturalidade – e, ainda, o seu tratamento dos termos teóricos.

Palavras-chave: identidade mente-corpo, interpretação radical, conteúdo restrito e lato, realismo modal, objetos possíveis.

Abstract

My purpose in this dissertation is to present the various theses defended by David Lewis concerning the ontology of mind and the proper treatment of mental content. Among those theses are his own variety of mind-body identity, his understanding of radical interpretation, his defense of the existence of a class of irreducibly egocentric (or first-personal) intentional attitudes and the relevance of narrow content to his overall theory of mind. To complete the task, I will also have to present some of D. Lewis's metaphysical ideas that underlie various of his proposals in philosophy of mind – namely, his genuine modal realism and the classification of properties by their degree of naturalness – and, finally, his treatment of theoretical terms.

Keywords: mind-body identity, radical interpretation, narrow and broad content, modal realism, possible objects.

Introdução.

David Kellogg Lewis (Oberlin, Ohio, 28 de setembro de 1941 – Princeton, New Jersey, 14 de outubro de 2001) foi um importante filósofo norte-americano, certamente um dos mais influentes e originais da sua geração. Frank Jackson e Graham Priest (2004: 1), por exemplo, escreveram que «[…] Lewis has an undeniable place in the history of analytical philosophy. His work defines much of the current agenda in metaphysics, philosophical logic, and the philosophy of mind and language». Com apenas um pouco mais de detalhe, pode-se dizer que o trabalho de D. Lewis acerca da modalidade, da causalidade, das leis da natureza, ou acerca da noção de propriedade e do uso do conceito de identidade em várias áreas da filosofia contribuiu enormemente para a respeitabilidade e centralidade que a metafísica foi readquirindo durante a segunda metade do século vinte no contexto da tradição analítica. Embora menos regularmente, D. Lewis também escreveu sobre epistemologia, filosofia da ciência, teoria dos valores, filosofia da religião, filosofia da matemática, teoria da decisão e semântica formal.

Orientado por W. V. Quine, D. Lewis fez o doutoramento na Universidade de Harvard com uma tese que veio a ser editada como *Convention* (1969a). Devido a algumas diferenças marcantes nas abordagens de cada um, é raramente reconhecido que as ideias e o estilo de prosa de W. V. Quine permaneceram uma importante influência no trabalho de D. Lewis (Lewis 2015: 10-1). Entre 1966 e 1970 deu aulas na Universidade da Califórnia, Los Angeles (UCLA). Durante esse tempo, desenvolveu o seu interesse pela semântica formal através dos diálogos que manteve com Richard Montague, Barbara Hall Partee, Hans Kamp e David Kaplan. Em 1970 mudou-se para a Universidade de Princeton, onde se manteve até à sua morte em 2001.

Mesmo um resumo da biografia intelectual de D. Lewis estaria incompleto se não mencionasse a forte proximidade que ele manteve com a Austrália e os filósofos australianos. A amizade com J. J. C. Smart e David Armstrong, dois filósofos com quem partilhava uma abordagem semelhante em filosofia da mente, foi crucial nesta aproximação. A primeira visita que fez à Austrália aconteceu no verão de 1971, após ter sido convidado por J. J. C. Smart para dar algumas palestras na Universidade de Adelaide. A segunda visita foi em 1976 e Stephanie Lewis, a sua esposa, conta-nos que desde essa

altura até ao ano da sua morte D. Lewis foi à Austrália mais de vinte vezes, visitando os departamentos de filosofia das várias universidades australianas e participando assiduamente na conferência da Australasian Association of Philosophy (Lewis 2015: 3). A proximidade de D. Lewis à Austrália é tão marcante que muitos filósofos australianos o consideram um "australiano honorário" (Nolan 2005: 4). É provável até que muitas pessoas pensem que D. Lewis é australiano.

É importante referir as visitas de D. Lewis à Austrália porque tiveram impacto tanto a nível pessoal como filosófico. Daniel Nolan explica assim as razões para D. Lewis ter encontrado na Austrália muitos colegas com quem teve imensas discussões frutíferas acerca de tópicos de interesse comum:

> «Lewis shared many Australian philosopher's preference for a down-to-earth, no-nonsense writing style, and many of Lewis's philosophical proclivities were shared by prominent Australian philosophers. A taste for philosophical materialism, a respect for the natural sciences, and an unabashed sense that metaphysical problems are real questions whose answers we can make real progress towards answering are only three of the philosophical traits Lewis shared with many Australian philosophers.» (Nolan 2005: 4)

No que me diz respeito, cabe-me dizer que o meu interesse na filosofia de D. Lewis surgiu há aproximadamente dois anos, enquanto frequentava o curso de Filosofia da Mente que está integrado no plano de estudos deste mestrado. Na altura, estava a tentar encontrar um tema para tratar no trabalho final e debrucei-me mais aprofundadamente sobre o argumento que Saul Kripke apresenta em *Naming and Necessity* (1980) contra a possibilidade da teoria da identidade entre a mente e o corpo. Esse argumento parte da ideia de que a identidade é necessária para concluir que um estado mental não pode ser idêntico a um estado físico, tendo em conta que é possível termos um deles sem termos o outro. Assumindo que a dor é idêntica ao disparo de fibras-C, então, defende S. Kripke, a dor é necessariamente idêntica ao disparo de fibras-C. Em qualquer mundo possível, qualquer criatura que encontremos a sentir dor tem um sistema nervoso em que ocorre um disparo de fibras-C, e vice-versa. Como isto é absurdo, chegamos à conclusão de que há alguma coisa de errado com a ideia de que os estados mentais são idênticos a estados físicos.

Na altura não fiquei convencido com este argumento. Pareceu-me plausível a ideia de que a identidade é necessária, apesar de hoje achar que essa questão é particularmente complexa.[1] Mas o que me pareceu, como ainda hoje me parece, foi que era pouco segura a suposição, em que o argumento se baseia, de que os termos mentais são designadores rígidos. É igualmente provável, pensei, e ainda continuo a pensar, que estes termos sejam equivalentes a descrições definidas que denotam em cada mundo possível aquilo, seja o que for, que nesse mundo exemplifica certas propriedades «topicamente neutras», para utilizar as palavras de J. J. C. Smart (1959: 150). Podemos assim manter a necessidade da identidade considerando ao mesmo tempo que é verdade de uma maneira contingente que os estados mentais são idênticos a estados físicos. Ainda que uma coisa seja idêntica a si própria em todos os mundos possíveis, um enunciado de identidade pode ser contingente desde que um dos termos que nele ocorre não seja um designador rígido. Um exemplo simples serve para ilustrar esta ideia. Enquanto que 'George Washington' é um designador rígido e denota George Washington em todos os mundos possíveis (em que ele existe), 'o primeiro presidente dos EUA' é uma descrição definida que denota em cada mundo possível aquela pessoa que nesse mundo é o primeiro presidente dos EUA. George Washington é necessariamente idêntico a George Washington, mas como o primeiro presidente dos EUA poderia ter sido outra pessoa, não é necessariamente verdade que George Washington é idêntico ao primeiro presidente dos EUA, apesar de isso ser verdade no mundo atual. Ora, se termos mentais como 'dor', 'experiência de vermelho' e 'crença de que está a chover' não são designadores rígidos, podemos defender a verdade contingente da teoria da identidade.

A teoria da identidade, para D. Lewis, não é uma hipótese que deve ser defendida pela sua parcimónia ontológica. Em vez disso, é uma teoria que, segundo ele, é implicada por certas suposições que considera plausíveis acerca do mundo atual. Mais concretamente, D. Lewis defende que a identidade entre a mente e o corpo se segue de uma conceção materialista do mundo, ou até quase-materialista, em que apenas eventos físicos têm impacto nos restantes eventos físicos – e todos os restantes eventos são no máximo epifenómenos, ou estão completamente desconectados da realidade física. D.

1 Creio que D. Lewis também considera complexa esta questão (*ver* secção 1.5.3).

Lewis acredita que uma descrição completa do mundo atual necessita apenas de referir propriedades e relações próximas daquelas que são referidas nas teorias físicas contemporâneas, e que provavelmente serão descobertas no desenvolvimento destas teorias. O mundo atual, para D. Lewis, consiste apenas num mosaico de propriedades físicas instanciadas por pequenas partículas arranjadas em certas relações espaciotemporais. Tudo o resto – a causalidade, as essências, a mente e a linguagem, entre outras coisas – é superveniente a este mosaico. Esta conceção materialista é marcante em muitas das propostas de D. Lewis em metafísica e filosofia da mente.

A teoria da identidade segue-se de uma conceção materialista ou quase-materialista do mundo atual porque D. Lewis considera que os termos mentais devem ser analisados como termos que denotam os estados que integram uma rede causal complexa que serve como intermediário entre os estímulos e os comportamentos das pessoas. Qualquer que seja o estado que ocupa um determinado papel nessa rede causal será o referente do termo mental correspondente. Assumindo que todos os papeis causais desse género são ocupados no mundo atual por estados físicos, é óbvio que todos os termos mentais terão estados físicos como referentes (se tiverem algum referente).

O significado dos termos mentais, de acordo com D. Lewis, tem origem na psicologia popular – uma teoria implicitamente utilizada por todos os humanos para prever e explicar o comportamento de cada um. É o conjunto de trivialidades que fazem parte dessa teoria que D. Lewis pensa que contém o significado dos vários termos mentais. A análise desses termos segue as linhas gerais do método de definição que D. Lewis propõe para os termos teóricos em geral. Suponhamos que um termo é introduzido numa teoria científica e que não foi alguma vez usado fora do contexto dessa teoria. Então, propõe D. Lewis, o significado desse termo é dado pela teoria e, acrescenta, podemos alcançar uma definição explícita através de alguns mecanismos formais.

A parte da psicologia popular acerca das crenças e desejos, de acordo com D. Lewis, tem a ver com a racionalidade dos agentes, contendo generalizações que afirmam que escolhemos o comportamento que serve da melhor maneira os nossos desejos de acordo com as nossas crenças, e que vamos alterando as nossas crenças e os nossos desejos instrumentais tendo em conta a evidência a que temos acesso através dos

estímulos que recebemos do ambiente externo. Esta é a base da teoria do conteúdo defendida por D. Lewis. A ideia é que o conteúdo das nossas crenças e desejos é aquele que permite descrever-nos como sendo em larga medida racionais. A certa altura, no entanto, D. Lewis vai ter de expandir esta teoria, porque diferentes interpretações permitem racionalizar adequadamente as mesmas disposições de comportamento e de ajuste das crenças aos estímulos. Para resolver este problema, D. Lewis vai restringir as interpretações possíveis àquelas que contêm propriedades razoavelmente naturais. Entre atribuir a crença de que as esmeraldas são verdes ou a de que as esmeraldas são verduis devemos preferir a primeira, porque categorizar alguma coisa como verdul é menos natural do que categorizá-la como verde. Esta classificação das propriedades em mais e menos naturais é uma importante tese metafísica de D. Lewis que ele considera ser útil na formulação de diferentes teorias.

Esta teoria implica a importante tese de que o conteúdo das crenças e desejos e o papel causal dos estados mentais não variam independentemente um do outro, o que significa que aquilo que se passa no ambiente externo a um sujeito acaba por ser irrelevante para o conteúdo desses estados. A famosa experiência de pensamento da Terra Gémea (Putnam 1973, 1975), e outras semelhantes, tentam demonstrar exatamente o contrário. A ideia que se retira destas experiências é que certas atribuições de conteúdo que fazemos na linguagem comum têm um valor de verdade que depende parcialmente daquilo que existe em torno do sujeito. D. Lewis tenta acomodar estas atribuições afirmando que elas não são apenas acerca da atividade psicológica dos sujeitos. O conteúdo que atribuem é derivado de um conteúdo mais básico e psicologicamente mais relevante.

D. Lewis aceita que os conteúdos dos vários estados mentais são objetos. Mas que tipo de objetos, exatamente? Até certa altura, ele diria que eram proposições – entidades abstratas com um valor de verdade que depende unicamente da maneira como o mundo é. As proposições parecem objetos adequados para as crenças e os desejos se assumirmos que o conteúdo desses estados mentais é apenas acerca do mundo. Mas, a certa altura, D. Lewis passa a defender que o conteúdo de alguns desses estados é irredutivelmente acerca do sujeito. Em vez de proposições, D. Lewis propõe que pensemos nos objetos dessas

crenças e desejos como propriedades. Vai até mais longe e defende que podemos utilizar propriedades mesmo quando as proposições são adequadas.

As propriedades são concebidas por D. Lewis como conjuntos e a relação de instanciação com a relação de pertença. Dizer que uma flor instancia a propriedade de ser vermelha é dizer que essa flor pertence ao conjunto de coisas vermelhas. É conhecida a objeção de que identificar propriedades com conjuntos leva à identificação de diferentes propriedades acidentalmente coextensivas, como as de ter um coração e de ter rins. Mas D. Lewis escapa a esta objeção utilizando os recursos oferecidos pela famosa – e, para muita gente, extravagante – tese de que os mundos meramente possíveis e os seus habitantes existem e são entidades do mesmo tipo que o mundo atual e nós que nele habitamos. A ontologia de D. Lewis está por isso povoada de objetos meramente possíveis, e as propriedades são identificadas com conjuntos de coisas espalhadas por todos os mundos possíveis, não apenas as que se limitam às fronteiras do mundo atual. Os membros atuais das propriedades acidentalmente coextensivas são os mesmos, mas há outras coisas, meramente possíveis, que evitam a identificação dessas propriedades.

Ora, este último ponto permite-me precisamente passar à apresentação das divisões do meu trabalho. Tendo em conta a importância da tese da pluralidade de mundos, ou realismo modal, para a ontologia que é usada na teoria da mente de D. Lewis, considerei relevante dedicar o primeiro capítulo, de longe o mais longo, a essa tese. Começo assim por clarificar como D. Lewis concebe os mundos possíveis e, de seguida, exponho o argumento que utiliza para defender a existência dos mundos assim concebidos. O argumento é essencialmente que assumir o realismo modal é útil para a teorização filosófica em diversas áreas, incluindo na abordagem à mente e ao conteúdo. Exponho, por isso, imediatamente a seguir, algumas das aplicações que D. Lewis faz dos mundos e dos objetos possíveis.

No segundo capítulo pretendo apresentar de uma maneira informal o método proposto por D. Lewis para a definição dos termos teóricos em geral, que será aplicado de uma forma crucial na análise do vocabulário mental. Essa aplicação é exposta imediatamente depois, e precede a apresentação da variedade de teoria da identidade entre a mente e o corpo defendida por D. Lewis.

O terceiro capítulo é uma continuação do segundo, e é dedicado à abordagem de D. Lewis aos estados mentais com conteúdo. Neste mesmo capítulo, entro novamente em território metafísico, apresentando a importante ideia de que as propriedades devem ser distinguidas pelo seu grau de naturalidade. No final do capítulo dedico-me brevemente ao conteúdo linguístico.

No quarto, e último, capítulo desenvolvo mais alguns temas relativos ao conteúdo mental. Mais concretamente, vou apresentar os argumentos de D. Lewis a favor da existência de uma classe de atitudes irredutivelmente egocêntricas – acerca do sujeito – e à ideia das propriedades como objetos das atitudes. De seguida, debruço-me finalmente sobre o carácter internalista da teoria do conteúdo mental de D. Lewis e como ele tenta conjugar, dentro dos recursos dessa teoria, as atribuições verdadeiras de conteúdo que dependem parcialmente daquilo que ocorre no ambiente externo ao sujeito.

Por falta de espaço, algum material teve de ser colocado em anexo. Esse material, creio, é um importante complemento ao que é dito no corpo da dissertação, acrescentando-lhe profundidade e rigor, apesar de não me parecer absolutamente indispensável. No primeiro anexo faço uma exposição formal do método de definição dos termos teóricos que é apresentado informalmente no segundo capítulo. No segundo anexo apresento brevemente a resposta de D. Lewis ao argumento do conhecimento de Frank Jackson contra o materialismo, o que complementa aquilo que é exposto ao longo do segundo capítulo. No terceiro anexo apresento as várias teorias metafísicas que D. Lewis considera viáveis para explicar a naturalidade das propriedades. E, por fim, no quarto anexo exponho os contornos do tipo de gramáticas que D. Lewis apresenta em "General Semantics" (1970b). Um contacto, mesmo que superficial, com essa conceção de gramática pode ser relevante na compreensão daquilo que é dito acerca do conteúdo linguístico no primeiro e terceiro capítulos.

Esta dissertação cobre um número aparentemente excessivo de temas. Apesar de estar centrada em torno da teoria da mente e do conteúdo, muitas páginas são dedicadas a questões sobre metafísica e linguagem. Em minha defesa, porém, deixo esta confissão de D. Lewis:

> «I should have liked to be a piecemeal, unsystematic philosopher, offering independent proposals on a variety of topics. It was not to be. I succumbed too often to the temptation to presuppose my views on one topic when writing on another. Most notably, my realism toward unactualized possibles shows up in nearly every paper in the book.» (Lewis 1983c: ix)

A sistematicidade da filosofia de D. Lewis pode exigir que a maneira como trata certos tópicos não seja abordada sem que haja uma familiaridade com a maneira como trata outros tópicos. Estou certo de que nem todas as partes da filosofia de D. Lewis dependem totalmente de outras. Considero até que na maior parte dos casos a conexão entre as partes surge mais pela forma como D. Lewis escolhe formular os problemas e as suas teses do que pelo conteúdo das mesmas. Mas o trabalho de distinguir o que depende, e o que não depende, de quê é, neste contexto, extremamente complicado. Eu sei que a teoria do conteúdo de D. Lewis depende da tese de que existem propriedades mais naturais que outras. Mas não tenho a certeza, por exemplo, de que o realismo modal e a teoria das propriedades como conjuntos de objetos possíveis sejam premissas indispensáveis para a conclusão a que chega D. Lewis de que as proposições não podem ser o objeto de todas as atitudes. Ainda assim, a apresentação dos passos através dos quais D. Lewis chegou a ela é impossível de ser compreendida sem falar do realismo modal e da teoria das propriedades como conjuntos de objetos possíveis.

.

Capítulo 1 – A pluralidade de mundos

1.1 – A vasta ontologia dos mundos

As preocupações filosóficas de David Lewis e de Willard Van Orman Quine são, em vários aspetos, muito diferentes. Como veremos, D. Lewis tem um enorme respeito pelo senso-comum e, por essa razão, vê como inaceitável a ideia de abandonar as partes do nosso discurso popular que envolvem, entre outras coisas, conceitos modais e atribuições de atitudes intencionais de modo a obter uma teoria do mundo mais económica e logicamente menos problemática, como é proposto por W. V. Quine (Quine 1960: cap. 6). Ainda assim, as abordagens de cada um deles têm em comum alguns pontos que são extremamente relevantes. A um nível fundamental, as teorias do mundo de D. Lewis e W. V. Quine são mais semelhantes do que aparentam ser, e isso reflete-se, por exemplo, nas categorias de entidades que aparecem nas ontologias a que cada um deles adere. Como é dito por Gideon Rosen, falando mais especificamente acerca da ontologia de D. Lewis:

> «Viewed from a certain distance, David Lewis's ontological scheme is simplicity itself. Absolutely everything that exists, according to Lewis, is either a spatiotemporal particular, or a set theoretic construction from such particulars, or a mereological aggregate of such items. Entities that other writers treat as *sui generis* – properties, relations, events, propositions, possible worlds and individuals, mental contents, languages, linguistic meanings – are all either identified with concrete things […] or with set theoretic constructions therefrom. Lewis inherited this scheme from Quine. Lewis and Quine disagree about the inventory of particulars, and perhaps about the principles of set theoretic construction. But they agree about this much: when God made the world(s), he made the concrete individuals; he laid down the principles of set theory, and then he stopped.» (Rosen 2015: 382)

Apesar de haver diferenças noutros aspetos, tanto a ontologia de D. Lewis como a de W. V. Quine apresentam uma simplicidade incrível. Para ambos os autores, existem apenas particulares concretos – particulares espaciotemporais e somas mereológicas de cada pluralidade destes – e a hierarquia de conjuntos e classes construídos a partir deles.[2]

[2] O próprio D. Lewis, no entanto, não apresentaria desta maneira o inventário das categorias de entidades presentes na sua ontologia, porque duvida que haja alguma coisa que toda a gente quer dizer

Essa simplicidade é ainda mais incrível no caso do esquema ontológico de D. Lewis, tendo em conta que, enquanto a ontologia de W. V. Quine é um ambiente hostil ao discurso modal e psicológico, D. Lewis encontra uma maneira de analisar esses idiomas através da identificação de entidades banidas do universo de W. V. Quine, como mundos e indivíduos possíveis, propriedades, relações, proposições, conteúdos mentais e significados linguísticos, com particulares concretos ou conjuntos, mais ou menos complexos. Daniel Nolan afirma que «Lewis is […] inclined to think that concrete objects plus mathematical objects such as sets are all we need to explain reality. Aspects of reality such as meaning, morality or necessity are to be explained ultimately in these terms.» (Nolan 2005: 203) É verdade que W. V. Quine concorda com a ideia de que a realidade pode ser explicada inteiramente através do tipo de entidades que ele e D. Lewis reconhecem, mas com a consequência – ou o custo, de um ponto de vista lewisiano – de abdicar de conceber a modalidade e as atitudes intencionais, assim como tudo aquilo que destas é consequência, como aspetos da realidade.

É crucial, neste contexto, o *realismo modal* adotado por D. Lewis: a tese de que os mundos possíveis existem e são entidades concretas do mesmo tipo que o mundo de

quando fala de entidades concretas e abstratas, considerando que há várias maneiras de explicar essas noções (Lewis 1986b).

que fazemos parte (Lewis 1973b: 84-5, 1986b: 1-3).[3, 4] Assim como o mundo atual é o lugar onde estamos nós e as coisas que nos rodeiam, de acordo com D. Lewis os restantes mundos são outros lugares, que diferem do atual por terem outros habitantes, numericamente – e, muitos deles, qualitativamente[5] – distintos dos habitantes do mundo

[3] D. Lewis não considera útil caracterizar uma entidade como concreta, duvidando de que exista alguma coisa que toda a gente queira dizer quando caracteriza assim uma entidade. Ele encontra várias maneiras – não equivalentes – de explicar essa noção, e é interessante ver como se classificam os mundos possíveis de acordo com cada uma delas. A primeira consiste em enumerar alguns exemplos paradigmáticos de entidades concretas e abstratas. Diz-se assim que são concretas coisas como protões, estrelas e gatos, e abstratas coisas como números. Desta forma podemos dizer que pelo menos algumas das partes dos outros mundos são concretas, já que essas partes vão ser precisamente coisas como protões, estrelas e gatos. Temos de permanecer silenciosos, ainda assim, acerca de hipotéticas partes de outros mundos como pontos espaciotemporais, universais e tropos, e também acerca dos mundos inteiros. Talvez alguns mundos bastante pequenos sejam idênticos a protões, estrelas e gatos, e são por isso considerados concretos nesta conceção, mas de outros, como o mundo atual, compostos por uma enorme diversidade de coisas, nada se pode dizer. A segunda maneira passa por identificar a distinção entre concreto e abstrato com a distinção entre outras categorias de entidades, como, por exemplo, a distinção entre indivíduos e conjuntos, ou entre particulares e universais, ou entre particulares individuais e quaisquer outras coisas. D. Lewis considera que os mundos são indivíduos particulares, por isso não há qualquer dúvida de que pensando assim na distinção entre concreto e abstrato, os mundos são concretos. A terceira maneira é a de dizer que as entidades abstratas são aquelas que não têm uma localização espáciotemporal, não estabelecem relações causais com outras coisas, e duas delas nunca são indiscerníveis, dizendo-se de seguida que as entidades concretas são aquelas que não são abstratas. De acordo com D. Lewis, não existe um espaço-tempo unificado em que todos os mundos se localizam, e entre eles não há interação causal. No entanto, dentro de cada mundo, as suas partes estão certamente localizadas e interagem entre si, e além disso algumas delas são indiscerníveis. Por isso, partes dos mundos são concretas. Os mundos são agregados das suas partes, que são concretas, por isso intuitivamente também devem ser considerados concretos nesta conceção, apesar de, rigorosamente falando, não terem qualquer localização nem participação em nexos causais. A quarta, e última, maneira é a de considerar que as entidades abstratas são abstrações das coisas concretas, tendo por isso uma menor especificidade. Para D. Lewis, os mundos são claramente concretos, se isso for uma maneira de dizer que não são uma abstração de alguma outra coisa: os mundos são completamente específicos (Lewis 1986b: 82-6).

[4] Nesta conceção de mundos possíveis fica excluída a hipótese de se dizer, por exemplo, que o mundo atual é concreto e os restantes são abstratos, que o mundo atual é um indivíduo e os restantes são propriedades, universais, ou conjuntos, ou que o mundo atual é uma entidade existente e os restantes são uma exótica espécie de entidades não-existentes, entre outras alternativas deste género. D. Lewis também pretende excluir a hipótese, que considera ininteligível, de o mundo atual distinguir-se dos restantes pela sua maneira de existir, como se o mundo atual e os seus habitantes existissem de uma maneira mais real ou mais perfeita que os mundos meramente possíveis, que existiriam de uma maneira mais sombria ou esbatida (Lewis 1986b: 2-3).

[5] Para D. Lewis, como veremos mais à frente, nenhuma coisa faz parte de mais do que um mundo. A haver universais, estes seriam a única exceção a esta generalização, mas estes objetos não chegam a ser aceites por D. Lewis na sua ontologia, apesar de também não serem abertamente rejeitados (*ver* anexo 3). Por isso, quaisquer partes de diferentes mundos são – ignorando a complicação criada pelos universais – numericamente distintas. Mas, se houver mundos possíveis indiscerníveis, diferentes mundos contêm habitantes qualitativamente idênticos. D. Lewis é agnóstico quanto a esta hipótese (Lewis 1986b: 87). Em todo o caso, a sua aceitação ou rejeição não teria qualquer impacto significativo na sua teoria metafísica geral.

atual.[6] É a aceitação dos mundos possíveis, rejeitados por W. V. Quine (1953: 4), que permite a D. Lewis encontrar vários tipos de conjuntos que podem desempenhar o papel de propriedades, relações, proposições, conteúdos mentais e significados linguísticos, de maneiras que serão apresentadas mais à frente. Podemos dizer, deste modo, que a diferença mais importante entre as ontologias de D. Lewis e W. V. Quine é quantitativa: apesar de ambas serem igualmente simples no inventário das categorias de entidades existentes, a realidade é muito mais vasta e diversa de acordo com D. Lewis do que de acordo com W. V. Quine.

A tese de que existem vários mundos concretos implica a tese de que existem de igual modo os habitantes desses mundos. Habitar um mundo é basicamente (1) ser um indivíduo que faz parte desse mundo ou (2) ser um conjunto que está localizado nesse mundo. (Esta bifurcação entre os indivíduos e os conjuntos exige um comentário. Apenas indivíduos podem ser partes de um indivíduo, de acordo com D. Lewis. Por esse motivo, um mundo, sendo um indivíduo, não tem qualquer conjunto como parte. Ainda assim, D. Lewis está inclinado a aceitar que um conjunto pode estar localizado onde se encontram os seus membros (Lewis 1986b: 94-5) e, desse modo, mesmo não fazendo parte de um mundo, um conjunto de indivíduos que fazem parte de um mesmo mundo encontra-se exatamente nesse mundo, assim como o conjunto que tem como único elemento esse conjunto, e por aí em diante.) Às partes de cada mundo, próprias e impróprias, chamamos indivíduos *possíveis*. (Um mundo é um indivíduo possível que é uma parte imprópria dele mesmo.) Através de um princípio irrestrito de composição (Lewis 1983c: 39, 1986b: 211-

[6] Falar dos mundos como lugares pode levar ao erro de se pensar que D. Lewis concebe os mundos como coisas parecidas com regiões no espaço-tempo, por exemplo, que podem ou não estar ocupadas, ou com garrafas que podem ou não conter cerveja. Nesta conceção, se considerarmos um mundo que contém como seu único habitante uma árvore (e as partes da árvore, claro, tendo em conta que a relação *ser parte de* é transitiva) estaríamos a considerar duas coisas diferentes: a árvore e o mundo que a contém. Existiria também a hipótese de haver um mundo que estivesse completamente vazio, que não contivesse nada dentro dele. Mas a conceção de D. Lewis, pelo contrário, é a de que uma árvore sozinha num mundo é ela própria esse mundo. Não há nada mais num mundo que os seus habitantes, por isso é que a árvore é idêntica ao mundo em que vive, e nenhum mundo pode ser desabitado. Utilizando os conceitos da mereologia, dizemos que um mundo é a soma dos seus habitantes, a coisa menos inclusiva (i. e., com menos partes) que inclui todos os seus habitantes como partes (Lewis 1986b: 69). Uma das consequências do realismo modal de D. Lewis é, assim, a de que não é possível não existir nada. Esta consequência pode ser matéria para uma objeção ao realismo modal.

13),[7] a existência dos vários mundos concretos implica ainda a existência de indivíduos que têm partes diferentes em vários mundos. Estes chamam-se indivíduos *impossíveis*, porque não há maneira de eles poderem ser atuais: se qualquer outro mundo fosse o atual, na melhor das hipóteses apenas uma parte própria desses indivíduos seria atual. Enquanto que um indivíduo possível está *integralmente* num mundo, um indivíduo impossível está *parcialmente* em vários. Devemos estender esta categorização aos conjuntos. Enquanto que alguns conjuntos têm apenas elementos que habitam um único mundo, e por isso estão integralmente nesse mundo, alguns outros têm elementos em vários mundos, e por isso estão parcialmente em vários deles. Existem ainda objetos na ontologia de D. Lewis que não estão integralmente nem parcialmente em qualquer mundo – mais concretamente, os conjuntos puros, que não são construídos a partir de indivíduos. (Se quisermos identificar os números com conjuntos puros, então os números também não estão em qualquer mundo.) Há ainda que considerar uma outra maneira de estar num mundo – a de *existir a partir do ponto de vista* de um mundo:

> «When we evaluate the truth of a quantified sentence, we usually restrict the domain and quantify over less than all there is. If we evaluate a quantification at a world, we will normally omit many things not in that world, for instance the possible individuals that inhabit other worlds. But we will not omit the numbers, or some of the other sets. Let us say that an individual exists *from the standpoint of* a world iff it belongs to the least restricted domain that is normally – modal metaphysics being deemed abnormal – appropriate in evaluating the truth at that world of quantifications. I suppose that this domain will include all the individuals in that world; none of the other individuals; and some, but not all, of the sets. There will be many sets that even exist from the standpoint of all worlds, for instance the numbers. Others may not; for instance the unit set of a possible individual might only exist from the standpoint that the individual is in.» (Lewis 1983c: 40)

7 «I claim that mereological composition is unrestricted: any old class of things has a mereological sum. Whenever there are some things, no matter how disparate and unrelated, there is something composed of just those things. Even a class of things out of different worlds has a mereological sum.» (Lewis 1986b: 211) Este princípio implica que, para além das somas mereológicas de átomos idênticas a uma barra de ouro e a um gafanhoto, existe ainda a soma da barra de ouro e do gafanhoto, mesmo que a barra esteja em Israel e o gafanhoto na Indonésia. Estes objetos estranhos que assim são obtidos devem, no entanto, ser ignorados nos nossos idiomas quotidianos.

Um proponente do realismo modal tem de admitir que a verdade de uma frase é normalmente concebida como relativa a cada mundo. Exceto quando pretendemos falar acerca do que é possível e necessário, avaliamos a verdade de uma frase apenas na pequena parte da realidade que é atual, ignorando tudo o que é meramente possível. É por isso que podemos dizer corretamente que não há unicórnios, apesar de o espaço lógico estar repleto dessas criaturas. A ideia é que implicitamente estamos a falar apenas do mundo atual, de uma maneira análoga àquela como às vezes falamos apenas acerca daquilo que está em nossa casa quando dizemos corretamente que não há mais cerveja. Normalmente, é verdade num mundo aquilo que é verdade quando o domínio de quantificação é restringido ao que existe (integralmente, na maior parte das vezes) nesse mundo (Lewis 1986b: 5-6). Em particular, é atualmente verdade aquilo que é verdade quando o domínio de quantificação inclui apenas as coisas que existem no mundo atual. (Esta restrição afeta não apenas os valores das variáveis quantificadas, mas também, entre outras coisas, o referente dos termos singulares e a extensão dos predicados.) Mas às vezes a avaliação adequada de uma frase acerca do que acontece num mundo exige que se alargue o domínio dos quantificadores a coisas que se encontram fora dele, e por causa do relevo que assim adquirem dizemos que existem desde o ponto de vista desse mundo.

1.2 – Atualidade e mera possibilidade

Uma teoria que aceite a existência de uma pluralidade de mundos e a atualidade de um deles deve ainda ter um capítulo que explique em que consiste a atualidade – que características tem um mundo de ter para ser o mundo atual? – ou pelo menos que esclareça que essa propriedade é primitiva e irredutível. (A mesma necessidade não aparece numa teoria que aceite que a realidade é formada apenas pelo mundo atual: a atualidade pode então ser identificada com a mera existência.) D. Lewis defende uma teoria indexical da atualidade, apresentada em "Anselm and Actuality" (1970a). Aí, escreve:

> «I suggest that "actual" and its cognates should be analyzed as *indexical* terms: whose reference varies, depending on relevant features of the context of utterance. The relevant feature of context, for the term "actual", is the world at which a given

> utterance occurs. According to the indexical analysis I propose, "actual" (in its primary sense) refers at any world w to the world w.» (Lewis 1970a: 184-85)

Um termo *indexical* (como 'eu', 'aqui', 'agora' e 'acima mencionado') tem um referente variável que depende de certas características relevantes do contexto em que está a ser utilizado. A ideia de D. Lewis é que 'atual' é um indexical que refere num certo contexto possível o mundo em que esse contexto está. O nosso mundo, assim, é o atual porque é aquele em que nos encontramos – ao falarmos do que é ou não atual, estamos localizados num certo contexto do espaço lógico que determina o nosso mundo como o referente de 'atual'. Do ponto de vista dos habitantes dos outros mundos, no entanto, é o mundo deles que é o atual, e o nosso passa assim a adquirir o estatuto de meramente possível. Não há nenhuma característica do mundo atual – ou de qualquer outro – que faça dele o referente adequado de 'atual'. A atualidade, por isso mesmo, é uma característica relativa.

D. Lewis considera que, para quem assume uma pluralidade de mundos, não há qualquer conceção adequada de atualidade a não ser como uma propriedade relativa. Em primeiro lugar, conceber a atualidade como absoluta impede-nos de explicar como é que temos a certeza, e é absurdo duvidar, de que vivemos no mundo atual:

> «The strongest evidence for the indexical analysis of actuality is that it explains why scepticism about our own world is absurd. How do we know that we are not the unactualized possible inhabitants of some unactualized possible world? We can give no evidence: whatever feature of our world we may mention, it is shared by other worlds that are not actual. Some unactualized grass is no less green, some unactualized dollars buy no less (unactualized) bread, some unactualized philosophers are no less sure they are actual. Either we know in some utterly mysterious way that we are actual; or we do not know it at all.» (Lewis 1970a: 186)

Em "Theories of Actuality" (1974), Robert Adams levanta a hipótese de que conhecemos a atualidade (absoluta, *ex hypothesi*) do nosso mundo através do apercebimento imediato da nossa própria atualidade:

> «[…] it can be maintained that actuality is a simple property which is possessed not only by the actual world as a whole, but by everything that exists in the actual world, and that we are as immediately acquainted with our own actuality as we

> are with our own thoughts, feelings, and sensations. It would be plausible, on this account of the matter, to suppose that acquaintance with our own actuality plays an important part in our acquisition of the concept of actuality, providing us with a paradigm of actuality, so that it would be reasonable to say, "If I am not actual, I do not know what actuality is."» (Adams 1974: 221)

Aqui, R. Adams assume que a atualidade não é apenas uma característica de um mundo inteiro, mas também dos habitantes desse mundo. É assim também na teoria de D. Lewis. Anteriormente foram introduzidos os conceitos de existir integralmente e existir parcialmente num mundo, e ainda de existir a partir do ponto de vista de um mundo. Tal como algumas coisas estão apenas parcialmente num mundo, também algumas coisas são apenas parcialmente atuais. E algumas outras, que não são atuais, nem sequer parcialmente, podem ser atuais *por cortesia*, ao existirem a partir do ponto de vista do nosso mundo (Lewis 1986b: 94-6).

Mas, de acordo com D. Lewis, mesmo que cada um de nós seja atual, não é aceitável dizer que temos um apercebimento desse estatuto que seja diferente de um modo relevante daquilo que acontece com os habitantes dos outros mundos, meramente possíveis: «[…] if Adams and I and all the other actual people really have this immediate acquaintance with absolute actuality, wouldn't my elder sister have had it too, if only I'd had an elder sister? So there she is, unactualised, off in some other world getting fooled by the very same evidence that is supposed to be giving me my knowledge.» (Lewis 1986b: 94) É impossível aquilo que não é verdade em nenhum mundo. Se as pessoas do mundo atual forem as únicas que têm um estado mental a que podemos chamar 'um apercebimento imediato da sua própria atualidade', então não é verdade em mais nenhum mundo para além do atual que as pessoas têm esse estado mental. É assim impossível que o mundo seja ligeiramente diferente de como efetivamente é e as pessoas continuem a ter consciência de que são atuais: isso não acontece em nenhum mundo. Bastava que uma borboleta tivesse movido as suas asas de uma outra maneira para que ocorresse a catástrofe de nenhum de nós ter qualquer conceção de que é atual. Esta consequência é absurda, e permite-nos dizer que habitantes de pelo menos alguns outros mundos têm uma experiência, ainda que ilusória, semelhante àquela que é o apercebimento da nossa

atualidade. Falar de um apercebimento imediato da atualidade não nos ajuda a afastar as potenciais dúvidas céticas relativamente ao nosso estatuto enquanto coisas atuais.

Além disso, temos que apenas a atualidade relativa permite acomodar o facto incontroverso de que é contingente qual dos mundos é o atual (Lewis 1986b: 94). Uma situação é contingente se variar entre mundos. É contingente que haja animais carnívoros se e só se em alguns mundos há animais carnívoros e noutros não. Da mesma maneira, é contingente que um certo mundo seja atual se e só se esse mundo for atual em alguns, mas não em todos os mundos. Assumindo que é contingente qual dos mundos é o atual, é óbvio então que a atualidade não é uma característica absoluta que separa um deles de todos os restantes, tendo em conta que nesse caso a questão da atualidade seria constante em qualquer mundo – e a contingência desapareceria.

1.3 – Individuação dos mundos

Quem assume que existe um único mundo pode dizer que 'mundo' significa 'a soma de tudo o que existe'. Mas essa ideia não pode ser partilhada por um proponente do realismo modal. Assumindo que há vários mundos, é necessário responder a esta questão: o que faz com que algumas das coisas existentes sejam mundos? Tendo em conta que D. Lewis concebe os mundos como somas de coisas, podemos reformular esta questão da seguinte maneira: o que faz com que uma soma de coisas seja um mundo, e não uma parte de um mundo, nem uma soma de vários mundos, nem uma soma de partes de vários mundos? Uma das formas possíveis de responder a isto é especificar que relação tem de existir entre várias coisas para que exatamente essas coisas e mais nenhumas pertençam a um mesmo mundo. Feito isto, dizer o que faz com que alguma coisa seja um mundo é bastante simples: um mundo é uma soma de coisas que se relacionam entre si de uma determinada maneira, e mais nenhuma outra coisa se relaciona dessa maneira com alguma dessas coisas. A estratégia que acabei de delinear é aquela que é seguida por D. Lewis:

> «[…] things are worldmates iff they are spatiotemporally related. A world is unified, then, by the spatiotemporal interrelation of its parts. There are no spatiotemporal relations across the boundary between one world and another; but

> no matter how we draw a boundary within a world, there will be spatiotemporal relations across it.» (Lewis 1986b: 71)

Com a ideia de que duas coisas pertencem ao mesmo mundo se e só se estiverem relacionadas espáciotemporalmente, chegamos a uma análise de 'mundo'. Dizer que uma coisa é um mundo é o mesmo que dizer que essa coisa é uma soma *máxima* de coisas que estabelecem entre si uma relação espáciotemporal.[8]

Há, no entanto, uma complicação que tem de ser tida em conta. D. Lewis nota que quando *nós* falamos em relações espaciotemporais estamos a falar de certas relações instanciadas no *nosso* mundo possível. De acordo com a teoria da relatividade essas relações são de uma determinada maneira, e de acordo com a mecânica clássica são de uma outra. A investigação empírica é favorável à primeira teoria, mas parece possível que a última fosse verdadeira. O que é possível é o que é verdade em algum mundo, o que significa que num mundo as relações espaciotemporais estão de acordo com a mecânica clássica. A pergunta que D. Lewis coloca é esta: serão as relações instanciadas no nosso mundo relativista e as instanciadas no mundo clássico as *mesmas*, apesar das diferenças que apresentam, ou serão *distintas*? Se forem as mesmas, nenhuma complicação tem de ser enfrentada. Mas, se forem distintas, deixamos de ter uma definição satisfatória da noção de 'mundos possíveis'. Isso acontece porque, nesse caso, alguns mundos não seriam unificados pelas relações espaciotemporais entre as suas partes, visto que nalguns deles a mecânica clássica é verdadeira e as *supostas* relações espaciotemporais descritas nessa teoria não são *verdadeiramente* relações espaciotemporais.

Perante esta situação, D. Lewis considera melhor modificar a sua proposta, e deixar esta questão em aberto, em vez de correr o risco de assumir sem justificação que as relações de que falam a teoria da relatividade e a mecânica clássica são as mesmas

---

8 Podemos ainda tornar a ideia mais explícita através destas três formulações ('R(*x*, *y*)' abrevia neste contexto '*x* e *y* estabelecem entre si uma relação espáciotemporal'): (1) *w* é um mundo se e só se para quaisquer *x* e *y* que são partes de *w* é verdade que R(*x*, *y*), e para nenhum *x* é verdade que exista um *y* que seja parte de *w* e R(*x*, *y*). (2) Assumindo que *x* é um objeto espáciotemporal se e só se para quaisquer *y* e *z* que sejam partes de *x* é verdade que R(*y*, *z*), temos então que *w* é um mundo se e só se *w* é um objeto espáciotemporal e a soma de *w* com qualquer *y* tal que *y* não seja uma parte de *w* não é um objeto espáciotemporal (van Inwagen 1986: 187). (3) *w* é um mundo se e só se há um conjunto não-vazio K tal que para quaisquer *x* e *y* que sejam membros de K é verdade que R(*x*, *y*), não existe nenhuns *x* e *y* tais que *x*, mas não *y*, seja um membro de K e R(*x*, *y*), e *w* é a soma de todos os membros de K (esta é, aproximadamente, a formulação que se encontra em (McDaniel 2004: 142).)

(Lewis 1986b: 75). A alternativa mais simples que considera é a de dizer que os mundos são unificados por relações *externas* naturais. Uma relação é externa se não for superveniente ao carácter intrínseco dos objetos relacionados, mas apenas ao composto deles, como acontece com as relações espaciotemporais. Esta alternativa acomoda a ideia de que outros mundos diferentes do atual não são unificados por relações espaciotemporais, mas por relações *parecidas* com estas, tendo conta que presumivelmente estas também são externas. Tem ainda a vantagem de ser uma abordagem mais geral, deixando em aberto a possibilidade de existirem mundos unificados por relações externas muito diferentes das espaciotemporais. Mesmo assim, D. Lewis não adota esta alternativa, apresentando uma hipótese em que se consideram certas relações como externas e naturais, e na qual a alternativa falharia. A hipótese é a de que as relações *ter a mesma carga* e *ter carga oposta* seriam relações externas (naturais), em vez de serem relações *internas* como normalmente as concebemos. A ideia comum é que essas relações são supervenientes ao carácter intrínseco das partículas que relacionam, sendo a parte relevante do carácter intrínseco das partículas neste caso a exemplificação das propriedades *ter carga positiva* ou *ter carga negativa*. Na hipótese que estamos a considerar, isso não poderia acontecer, e *ter carga positiva* e *ter carga negativa* teriam de ser analisadas de alguma maneira como propriedades extrínsecas das partículas. A pergunta agora a colocar é: assim compreendidas, as relações *ter a mesma carga* e *ter carga oposta* seriam instanciadas por pares de partículas em diferentes mundos? A resposta a esta questão é, como D. Lewis admite, aparentemente que sim, e por isso esta estranha hipótese mostra que a alternativa de unificar os mundos por quaisquer relações externas não é adequada (Lewis 1986b: 76-8).

A alternativa que D. Lewis vai adotar é dizer que os mundos podem ser unificados por relações espaciotemporais, mas também por relações *análogas* a estas. Para se perceber os contornos gerais da analogia que estaria em causa, deixo a passagem em que D. Lewis a apresenta:

> «At least some of the points of analogy should go as follows. (1) The relations are *natural*; they are not gruesome gerrymanders, not even mildly disjunctive. (2) They are *pervasive*: mostly, or perhaps without exception, when there is a chain of relations in the system running from one thing to another, then also there is a

> direct relation. (3) They are *discriminating*: it is at least possible, whether or not it happens at every world where the relations are present, that there be a great many interrelated things, no two of which are exactly alike with respect to their place in the structure of relations. (4) They are *external*: they do not supervene on the intrinsic natures of the *relata* taken separately, but only on the intrinsic character of the composite of the *relata*.» (Lewis 1986b: 75-6)

Uma das consequências de assumir que os mundos são unificados por relações espaciotemporais ou relações análogas a estas é que se está obrigado a dizer que não é possível existirem partes isoladas dentro de um único mundo. Nesta proposta, entre duas coisas que estão no mesmo mundo há sempre alguma distância espacial e temporal, por muito pequena ou grande que seja. (Duas partes isoladas estariam em dois mundos diferentes, ou seriam até elas próprias os dois mundos.) D. Lewis propõe algumas alternativas para tentar acomodar parcialmente aquilo que se tem em mente quando se pensa em mundos com partes desconectadas,[9] mas reconhece esta consequência como uma fraqueza da sua teoria. Apesar disso, vê esse custo como negociável, tendo em conta que a existência de mundos com a peculiaridade de não estarem unificados totalmente por relações espaciais e temporais (1) não é uma parte importante do nosso pensamento modal e (2) não é consequência de nenhum princípio geral interessante acerca do que é possível (Lewis 1986b: 71-3).

D. Lewis pensa que a única alternativa a aceitar que os mundos são unificados por relações espaciotemporais é adotar como primitiva a relação de pertença ao mesmo mundo, algo que acredita ser ainda menos aceitável (Lewis 1986b: 72). Ele não explica, no entanto, a razão de considerar que esta é a única alternativa, nem discute as vantagens e desvantagens que esta poderia ter relativamente à sua proposta. A mim parece-me que

---

9 As alternativas que propõe são estas: «There are at least four ways for one big world to contain many world-like parts. Each is a way that a world could be; and so, say I, each is a way that some world is. / (1) The spacetime of the big world might have an extra dimension. The world-like parts might then be spread out along this extra dimension like a stack of flatlands in three-space. / (2) The world-like parts might share a common spacetime. There might be several populations, interpenetrating without interaction in the single spacetime where all of them live. If so, of course the inhabitants had better not interact with the shape of their spacetime as we do with the shape of ours; else this interaction enables the different populations to interact indirectly with one another. / (3) Time might have the metric structure not of the real line, but rather of many copies of the real line laid end to end. We would have many different epochs, one after another. Yet each epoch would have infinite duration, no beginning, and no end. Inhabitants of different epochs would be spatiotemporally related, but their separation would be infinite. Or instead there might be infinitely many infinite regions laid out side by side in space; then there would have to be infinite spatial distances between points in different world-like regions.» (Lewis 1986b: 72)

não explicar em que consiste pertencer ao mesmo mundo tem a desvantagem de não nos explicar também como é que podemos ter a certeza de que Parménides e Marte, por exemplo, estão ambos no mesmo mundo – em geral, não nos permite compreender como é que podemos conhecer que dois objetos fazem parte do mesmo mundo. Talvez se pudesse dizer que é uma condição suficiente para pertencerem ao mesmo mundo que dois indivíduos estejam espáciotemporalmente relacionados, daí que possamos saber que Parménides e Marte habitam o mesmo lugar do espaço lógico. Mas agora o problema é este: com que legitimidade podemos dizer que a conexão espacial e temporal implica a pertença ao mesmo mundo? A proposta de D. Lewis, ao contrário desta hipótese, explica isto bastante bem: a conexão espacial e temporal implica a pertença ao mesmo mundo porque estar no mesmo mundo é exatamente isso – estar conectado espacial e temporalmente.

1.4 – Método em filosofia e razões para aceitar a pluralidade de mundos

D. Lewis acredita na existência de uma pluralidade de mundos concretos por uma razão puramente instrumental, considerando que essa hipótese é mais útil do que as hipóteses alternativas para a economia e simplicidade da teoria total do mundo que pretende construir.

Aceitar uma tese ontológica unicamente por causa da sua utilidade teórica é certamente uma coisa controversa, mas a ideia de que a utilidade de uma hipótese é uma boa razão para a aceitar como verdadeira faz sentido na perspetiva metodológica que D. Lewis favorece. Ele vê o aglomerado de opiniões comuns que a maior parte de nós partilha como formando uma teoria popular do mundo. De acordo com D. Lewis, o propósito do trabalho filosófico não deve ser o de substituir essa teoria por uma outra, criada *ex nihilo*, mas o de melhorar a teoria que aceitávamos previamente, tornando-a mais sistemática e económica (Lewis 1973b: 88). Ao ser levada a cabo esta tarefa, pode ser necessário abandonar algumas das nossas opiniões pré-existentes, mas essas modificações não devem ser profundas ao ponto de tornar a teoria inacreditável no nosso quotidiano. Para ser credível, uma teoria deve mostrar-se ao mesmo tempo sistemática e conforme àquilo que antes acreditávamos mais firmemente: este é o padrão que deve ser

tido em conta quando se escolhe aceitar uma teoria filosófica (Lewis 1986b: 133-5). Ainda assim, D. Lewis explica que há espaço em aberto para que os vários teóricos tenham opiniões diferentes relativamente à teoria que, de entre as alternativas, se adequa melhor a esse padrão:

> «Philosophical theories are never refuted conclusively. […] when all is said and done, and all the tricky arguments and distinctions and counterexamples have been discovered, presumably we will still face the question which prices are worth paying, which theories are on balance credible, which are the unacceptably counterintuitive consequences and which are the acceptably counterintuitive ones. On this question we may still differ. And if all is indeed said and done, there will be no hope of discovering still further arguments to settle our differences.
> It might be otherwise if, as some philosophers seem to think, we had a sharp line between "linguistic intuition," which must be unchallengeable evidence, and philosophical theory, which must at all costs fit this evidence. If that were so, conclusive refutations will be dismayingly abundant. But, whatever may be said for foundationalism in other subjects, this foundationalist theory of philosophical knowledge seems ill-founded in the extreme. Our "intuitions" are simply opinions; our philosophical theories are the same. Some are commonsensical, some are sophisticated; some are particular, some general; some are more firmly held, some less. But they are all opinions, and a reasonable goal for a philosopher is to bring them into equilibrium. Our common task is to find out what equilibria there are that can withstand examination, but it remains for each of us to come to rest at one or another of them.» (Lewis 1983c: ix)

As opiniões comuns são, para D. Lewis, o ponto de partida para a teorização filosófica, mas não constituem qualquer tipo de evidência inabalável a que as nossas teorias se devam conformar. É recomendável a atitude teoricamente conservadora, mas o afastamento da nossa teoria popular é um custo que pode ser compensado pelos benefícios trazidos pela simplicidade das nossas teorias. A avaliação de até que ponto o custo deve ser pago e de quando se torna inaceitável é uma questão que, ainda assim, não pode ser resolvida decisivamente através de argumentos e contra-argumentos: a certa altura os critérios de avaliação são subjetivos, e aquilo que conta para alguém como uma refutação de uma proposta pode parecer a outra pessoa o reconhecimento de alguma consequência negativa dessa mesma proposta, mas que deve ser aceite para que não sejam sacrificados alguns aspetos positivos que ela traz para a nossa teoria total do mundo. A uniformidade

em filosofia fica irremediavelmente posta em causa. E a procura de provas e demonstrações irrecusáveis também.

É por isso que D. Lewis não tem intenção, em *On the Plurality of Worlds* (1986b), de oferecer um argumento que demonstre a verdade do realismo modal ou que refute as teorias alternativas. O argumento que apresenta, pelo contrário, pretende mostrar (1) que o realismo modal oferece os recursos necessários para teorizarmos de uma maneira sistemática acerca de uma grande parte das nossas opiniões populares, (2) que essa tarefa é melhor sucedida que as alternativas em alguns aspetos importantes, e (3) que os custos que esta teoria tem devem por isso ser pagos.

D. Lewis compara a utilidade da pluralidade de mundos em filosofia com a utilidade que a ontologia da teoria dos conjuntos tem em matemática. O vocabulário da teoria dos conjuntos permite-nos definir as restantes noções matemáticas, e a partir dos seus axiomas podemos derivar todos os teoremas em várias áreas da matemática. O custo de reduzir toda a matemática à teoria dos conjuntos é a aceitação da vasta ontologia a que essa teoria nos compromete. É recomendável pagar este preço? D. Lewis considera que sim, e diz o mesmo acerca da aceitação do realismo modal: esta teoria oferece-nos a possibilidade de analisar redutivamente várias noções que à partida pareciam ter de ser tomadas como primitivas, e dessa maneira podemos minimizar o inventário de noções da nossa ideologia básica, aceitando uma ontologia mais vasta, ficando com uma teoria do mundo qualitativamente mais económica e simples.[10]

[10] Esta atitude de D. Lewis pode ser vista também como uma preferência por um princípio *qualitativo*, em vez de *quantitativo*, de parcimónia. Uma teoria é qualitativamente parcimoniosa se postular o mínimo de *tipos* diferentes de entidades; enquanto que é quantitativamente parcimoniosa se postular o mínimo de instâncias dos vários tipos. D. Lewis usa essa distinção em (1973b) para responder à objeção de que aceitar o seu realismo modal seria uma "ofensa" a um princípio como a navalha de Ockham. Contra isto, afirma: «I subscribe to the general view that qualitative parsimony is good in a philosophical or empirical hypothesis; but I recognize no presumption whatever in favor of quantitative parsimony. My realism about possible worlds is merely quantitatively, not qualitatively, unparsimonious. You believe in our actual world already; I ask you to believe in more things of that kind, not in things of some new kind.» (Lewis 1973b: 87) Apesar de D. Lewis estar a usar esta distinção para responder a uma objeção, e não a fazer uma apologia do seu realismo modal, esta adequa-se facilmente a este propósito. Aceitar uma teoria que postule a existência de inúmeros mundos como o nosso pode não ser quantitativamente parcimonioso, mas é-o qualitativamente. Qualquer teoria que pretenda manter intacto aquilo que se consegue analisar com a pluralidade de mundos terá de o fazer recorrendo a entidades de tipos diferentes que o nosso mundo atual – ou, pelo menos, terá de apelar a noções que deixam de ser analisáveis. Por causa disto, a alternativa será qualitativamente – ainda que não quantitativamente, já que deixa o nosso o mundo como o único existente – menos parcimoniosa que a hipótese da pluralidade de mundos.

Um dos problemas desta ontologia dos mundos possíveis é que desafia consideravelmente uma parte das nossas opiniões comuns. De acordo com o realismo modal são verdade muitas coisas estranhas e aparentemente inaceitáveis: existem muitos lugares reais aos quais nunca poderemos aceder, com infinitas pessoas parecidas connosco e outras coisas parecidas com aquilo que nos rodeia, e ainda para mais todas as ficções que podemos conceber – pelo menos as que não são impossíveis – existem algures, num dos mundos espalhados pelo espaço lógico, onde vamos encontrar unicórnios, dragões, duendes, cogumelos gigantes onde vivem humanos e plantas falantes, entre muitas mais fantasias. Se o senso comum não nega que isto aconteça, pelo menos recomenda acerca disso o mais profundo agnosticismo. Além disso, o realismo modal também nos diz que apenas uma parte pequena da realidade é atual. Assumindo que há uma ontologia a que estamos popularmente comprometidos, essa ontologia é *atualista*. A nossa conceção comum é a de que tudo o que existe, sem exceção, faz parte do mundo atual, que é identificado com a realidade total. (Esta conceção de atualidade é absoluta e não relativa ao contexto considerado.) Qualquer um de nós aceita que era possível que existissem coisas diferentes das atuais, certamente, mas acrescentamos que essas coisas não existem efetivamente. De acordo com o realismo modal, pelo contrário, dizer que era possível existirem coisas diferentes das atuais é o mesmo que dizer que há efetivamente, ainda que noutros mundos, coisas possíveis diferentes das atuais.

D. Lewis reconhece que o realismo modal ofende partes significativas da nossa teoria popular do mundo (Lewis 1986b: 133-5). Por isso, o seu argumento vai ter de envolver a comparação dos benefícios trazidos pelo realismo modal com aqueles que são trazidos pelas alternativas atualistas, que não exigem o pagamento de um custo tão elevado para as nossas opiniões pré-existentes. Em (Lewis 1986b: cap. 3), são consideradas três variedade de uma teoria chamada realismo modal *ersatz* (ou ainda *ersatzismo*). Esta teoria é atualista, mas não nega a existência de mundos e objetos possíveis, identificando-os (redutivamente) com entidades abstratas que fazem parte do inventário de coisas do mundo atual. (Cada variedade de ersatzismo difere das outras pelo tipo de entidades com que identifica os *possibilia*.) Falando de uma maneira rigorosa, um ersatzista não está a exprimir a mesma ideia que D. Lewis quando afirma que há mundos

possíveis. Ainda que ambos estejam a dizer que existem coisas deste ou daquele género que desempenham o papel teórico associado ao conceito de mundos possíveis por várias teorias em metafísica e semântica, por exemplo, há uma discordância profunda entre eles acerca do género de entidades que efetivamente desempenham esse papel.[11] Enquanto que D. Lewis fala de mundos genuínos, coisas que em nenhum aspeto relevante são diferentes do mundo em que vivemos, um ersatzista fala de representações abstratas do mundo e dos objetos concretos que servem para simular os mundos genuínos. É por isso que D. Lewis prefere chamar 'mundos (e objetos) possíveis ersatz' àquilo que um proponente do ersatzismo chama apenas 'mundos (e objetos) possíveis'.

Parte do argumento de D. Lewis é que nenhuma das variedades de ersatzismo consegue realmente oferecer os mesmos benefícios teóricos que o realismo modal genuíno. Cada uma das variedades tem as suas fraquezas e vantagens próprias, mas todas elas sofrem da falta de recursos que permitam uma análise redutiva da modalidade – uma tarefa que apenas a pluralidade de mundos permite levar a cabo, e sem a qual falha o projeto de criar uma teoria total do mundo mais económica, precisamente a razão pela qual D. Lewis pretende aceitar essa pluralidade. Este assunto será retomado mais à frente (*ver* secção 1.6), assim como algumas outras desvantagens específicas das variedades mágica e linguística de ersatzismo.

[11] Em "Two Concepts of Possible Worlds" (1986: 192-3), Peter van Inwagen sugere que o conceito de mundos possíveis deve ser concebido como um conceito funcional: precisamente como o conceito de coisas que desempenham um certo papel, determinado pelas teorias que falam acerca de mundos possíveis. Assumindo esta hipótese, podemos dizer que D. Lewis e os ersatzistas estão a dizer o mesmo quando afirmam que existem mundos possíveis, e que a discordância entre eles surge apenas quando o tópico da conversa é a metafísica da modalidade, i. e., quando se pretende explicar em que consistem os mundos possíveis. Peter van Inwagen compara esta situação às divergências entre um materialista e um dualista relativamente ao mental: ambos os teóricos estão a falar da mesma coisa (pessoas e os seus estados psicológicos), mas identificam essa coisa – um deles erroneamente, porque não é possível que ambos estejam certos neste aspeto – com coisas de diferentes géneros.

1.5 – A pluralidade de mundos aplicada

1.5.1 – Modalidade *de dicto*

Tendo em conta que o argumento de D. Lewis a favor da pluralidade de mundos passa por reconhecer a utilidade teórica desta hipótese, é conveniente neste contexto apresentar as aplicações dos mundos e objetos possíveis que D. Lewis vê como adequadas, começando pela aplicação ao âmbito da modalidade. A modalidade diz respeito aos conceitos de necessidade e possibilidade. É necessário aquilo que seria verdade mesmo que o mundo fosse de qualquer outra maneira como ele pode ser diferente daquela como efetivamente é – por exemplo, que nenhum solteiro é casado e que 5 + 7 = 12. É possível, por sua vez, aquilo que não é necessariamente falso. Tudo o que é verdade é possível, e algumas coisas são possíveis mesmo sendo falsas, como a existência de unicórnios. Aquilo que não é possível diz-se impossível. E, ainda, aquilo que nem é necessário nem impossível diz-se contingente. Em termos formais, usamos uma caixa (□) e um diamante (◇) antes de uma frase para expressar, respetivamente, a ideia de que ela é necessária e possível. Assim interpretados, a caixa e o diamante chamam-se operadores modais, e comportam-se de um modo semelhante aos quantificadores. Por exemplo, assim como '$\forall x\, \phi(x)$' é equivalente a '$\sim \exists x \sim\phi(x)$', também '$\Box\phi$' é equivalente a '$\sim \Diamond \sim\phi$'. Esta analogia pode ser levada ainda mais longe, e ser explicada, ao tratar-se mesmo os operadores modais como quantificadores. Tendo em conta que, de acordo com o realismo modal, cada maneira como o mundo pode ser é a maneira como um mundo efetivamente é, podemos dizer que é necessário aquilo que é verdade em todos os mundos, e é possível aquilo que é verdade em algum deles. (É impossível aquilo que não é verdade em nenhum, e contingente aquilo que é verdade nalguns, mas não em todos.) Os operadores de necessidade e possibilidade podem assim ser identificados, respetivamente, com um quantificador universal e existencial aplicado apenas a mundos possíveis. Consideremos estes exemplos:

(1) □(Nenhum solteiro é casado).

(2) ◇(Existem unicórnios).

Tendo em conta aquilo que acabamos de ver, (1) é equivalente a:

(1*) Todos os mundos são tais que em qualquer um deles nenhum solteiro é casado.

E (2) é equivalente a:

(2*) Existe um mundo em que existem unicórnios.

Como vimos antes (*ver* secção 1.1), é verdade num mundo (*ceteris paribus*) aquilo que é verdade quando se restringe o domínio dos quantificadores às coisas que existem a partir do ponto de vista desse mundo – ou seja, às coisas que existem integralmente ou parcialmente nesse mundo ou que são relevantes para avaliar a verdade de uma frase acerca desse mundo. O modificador 'em *w*' (sendo *w* um termo que denota um mundo qualquer) comporta-se da mesma maneira, como repara D. Lewis, que o modificador 'na Austrália', por exemplo (Lewis 1986b: 5). Esta evidente aproximação entre os mundos e lugares no espaço físico permite perceber como nesta teoria a modalidade, em vez de ser acerca de alternativas para a realidade inteira, diz antes respeito à consideração daquilo que acontece em várias regiões do espaço lógico, que é a realidade inteira. (Ainda assim, esta teoria não pretende negar que a modalidade diz respeito às alternativas meramente possíveis à atualidade: essas alternativas e a atualidade são apenas partes de uma realidade que as abrange a ambas.)

Por vezes, as afirmações modais estão restritas a mundos que estabelecem uma certa relação de *acessibilidade* com o mundo atual – i. e., que satisfazem uma certa condição, especificada a partir do ponto de vista do mundo em que nos encontramos. Quando quantificamos sem restrição sobre os mundos, estamos a falar daquilo que é *logicamente* (ou *metafisicamente*) necessário e possível. Mas existem modalidades restritas que constituem a interpretação correta de algumas expressões em certos contextos. Queremos, algumas vezes, falar daquilo que é apenas nomologicamente necessário e possível, e por isso restringimos a nossa atenção a mundos com as mesmas leis naturais que o mundo atual. É nomologicamente necessário, por exemplo, que o pão

alimente criaturas como nós, ainda que isso não seja logicamente necessário: isto quer dizer que em todos os mundos em que as leis da natureza são as mesmas que no mundo atual, mas não em todos os restantes, o pão alimenta criaturas como nós. Outras vezes queremos falar daquilo que é historicamente necessário e possível, e por isso falamos apenas dos mundos semelhantes ao nosso até ao momento presente. Outras vezes ainda queremos dizer que alguma coisa é obrigatória, permitida ou proibida, e nesse caso estamos a falar daquilo que ocorre, respetivamente, em todos, alguns ou nenhum mundo moralmente aceitável a partir do ponto de vista do mundo atual. Aqui estamos no âmbito da modalidade deôntica. Também podemos falar daquilo que é necessário ou possível de acordo com o conhecimento de um sujeito – aquilo que é epistemicamente necessário e possível. Nesse caso estamos a restringirmo-nos aos mundos acessíveis epistemicamente a esse sujeito.

1.5.2 – Modalidade *de re* (teoria das contrapartes)

Temos vindo até agora a considerar apenas a aplicação dos operadores modais a frases fechadas (i. e., que não contêm variáveis que ocorrem livres). Aplicados assim, os operadores de necessidade e possibilidade servem para formar afirmações de modalidade *de dicto* (acerca daquilo que é dito), que dizem como o mundo tem de ser ou como pode ser. Os operadores modais, ainda assim, podem também ser aplicados a frases abertas. Uma frase aberta expressa uma certa condição que pode ser satisfeita por alguns objetos (ou sequências de objetos, se houver mais do que uma variável livre), e serve para construir uma frase fechada através da introdução de quantificadores aplicados a cada uma das variáveis livres. Ao aplicar-se um operador modal a uma frase aberta forma-se uma outra frase aberta que é satisfeita por aqueles objetos que têm de satisfazer a frase original ou que podem satisfazer essa frase. Quando se introduz quantificadores para fechar uma frase modal aberta, estamos a afirmar de alguns objetos que eles têm certas características de uma maneira necessária ou possível. Neste caso, estamos a fazer afirmações de modalidade *de re* (acerca de uma coisa). Esta variedade de modalidade é normalmente usada para estabelecer uma distinção entre as características *essenciais* e *acidentais* de uma coisa. Uma característica faz parte da essência uma coisa quando essa

coisa tem necessariamente de a ter, e é um acidente de uma coisa quando essa coisa a pode ter, mas não necessariamente. Consideremos estes exemplos de operadores modais aplicados a frases abertas:

(3) □(*x* é um cão).

(4) ◇(*x* é cinzento).

Ao tratarmos os operadores modais como quantificadores, vamos entender os enunciados (3) e (4) do seguinte modo, respetivamente:

(3*) Em todos os mundos, *x* é um cão.

(4*) Em alguns mundos, *x* é cinzento.

Esta análise sugere que um objeto existe (integralmente) em vários mundos, e é dessa maneira que ele pode satisfazer as frases abertas '*x* é um cão' e '*x* é cinzento' não apenas no mundo atual, mas também em todos os outros mundos em que o objeto existe ou em alguns deles, de modo a satisfazer as frases (3*) e (4*) e, consequentemente, satisfazer (3) e (4). Mas esta análise contraria uma das ideias centrais da metafísica de D. Lewis, de acordo com a qual nenhum objeto existe (integralmente) em mais do que um

mundo.[12, 13, 14] Admitir essa hipótese levantaria aquilo a que D. Lewis chamou o problema dos *intrínsecos acidentais*. Este problema não se aplica à ideia de que há indivíduos que existem em vários mundos possíveis, desde que se acrescente que as qualidades intrínsecas desses indivíduos permanecem constantes nos vários mundos em que existem.

---

12 Existem objetos – aqueles que D. Lewis classifica como os objetos *impossíveis* (*ver* secção 1.1) – que têm partes em vários mundos (porque são uma fusão mereológica de objetos em diferentes mundos). Mas estes estão apenas *parcialmente* em vários mundos e não são relevantes para a consideração das propriedades modais dos objetos que existem no mundo atual. Por isso, na prática podem ser ignorados.

13 Em (Lewis 1968), D. Lewis afirma o seguinte: «Unactualized possibles, things in worlds other than the actual world have been deemed "entia non grata", largely because it is not clear when they are or are not identical. But identity literally understood is no problem for us. Within any one world, things of every category are individuated just as they are in the actual; things in different worlds are *never* identical […].» (Lewis 1968: 114) Em nota de rodapé, D. Lewis diz-nos que está a responder a esta passagem de W. V. Quine, em *Word and Object*: «In *possible* concrete objects, unactualized possibles, we have another category of doubtful objects whose doubtfulness can be laid to defective nouns, with as good reason at least as in the case of attributes and propositions. For here again, and more glaringly than in the case of intensions, there is perplexity over identity. Even when a position is specified, as in 'the possible new church on that corner', 'the possible hotel on the corner', the identity of position does not make the possible objects identical.» (Quine 1960: 245)

Uma outra famosa passagem em que W. V. Quine expressa o seu ceticismo relativamente aos *possibilia* é esta, de "On What There Is": «Take, for instance, the possible fat man in that doorway; and, again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? Are there more possible thin ones than fat ones? How many of them are alike? Or would their being alike make them one? Are no *two* possible things alike? Is this the same as saying that it is impossible for two things to be alike? Or, finally, is the concept of identity simply inapplicable to unactualized possibles? But what sense can be found in talking of entities which cannot meaningfully be said to be identical with themselves and distinct from one another? These elements are well-nigh incorrigible.» (Quine 1953: 4)

O problema que W. V. Quine vê nos *possibilia* é a obscuridade do seu critério de individuação: quando é que dois objetos possíveis são ou não idênticos? D. Lewis acredita que a restrição que impôs a esses objetos – a de que cada um existe unicamente no seu mundo – pode resolver as perplexidades quanto à sua individuação. Como afirma, ao existirem em cada um dos seus mundos, esses objetos são individuados, conforme a sua categoria, da mesma maneira que os objetos atuais, não havendo nada mais a dizer acerca da sua identidade noutros mundos. Desta forma, D. Lewis pretende escapar à necessidade de ter de apelar a critérios de individuação dos objetos possíveis que passem por apontar certas características que fazem referência ao mundo atual, como parece ser o caso em todas as tentativas de individuação que W. V. Quine apresentou e criticou.

14 Ainda que nenhum mundo tenha partes em comum – a não ser, talvez, universais – não existe qualquer problema, para D. Lewis, com a ideia de que duas coisas diferentes têm partes em comum. Um exemplo é o dos gémeos siameses que possuem uma mão em comum (Lewis 1986b: 199). Existem outros casos pouco problemáticos. Os pulmões, sendo uma parte do corpo, partilham com este todas as suas partes. A Escandinávia e a União Europeia têm partes em comum, e aqui neste caso nenhuma é uma parte da outra (pelo menos neste momento), como acontecia com os pulmões e o corpo. (Estou a assumir que 'União Europeia' é o nome de uma região do planeta. Se isto for inaceitável, substitua-se 'União Europeia' pela descrição 'a região ocupada pelos países que pertencem à União Europeia'.) E, como D. Lewis aceita um princípio irrestrito de composição mereológica (Lewis 1986b: 211-13), tem de aceitar que existe uma imensidão de casos deste género, e entre estes estarão casos em que uma parte de um mundo faz parte de um todo que tem partes em muitos outros mundos (nesse caso, essa parte de um mundo faz parte de um objeto impossível). Não é, por isso, este o motivo da rejeição da ideia de que vários mundos podem ter o mesmo indivíduo como parte.

Ele surge, no entanto, imediatamente quando se pretende que a identidade ao longo de vários mundos explique em que consiste para um indivíduo ter certas características de uma maneira essencial ou acidental. (A não ser que se pretenda defender a ideia, aparentemente absurda, de que todas as qualidades intrínsecas são essenciais.) Nesta leitura, é acidental a um indivíduo aquilo que *lhe* acontece em alguns mundos, mas não noutros. E provavelmente algumas das suas características acidentais dizem respeito à sua natureza intrínseca, o que significa que esta é variável nos diferentes mundos em que o indivíduo existe. Vejamos um exemplo em que este problema aparece:

> «Hubert Humphrey has a certain size and shape, and is composed of parts arranged in a certain way. His size and shape and composition are intrinsic to him. They are simply a matter of the way he is. They are not a matter of his relations to other things that surround him in this world. Thereby, they differ from his extrinsic properties such as being popular, being Vice-President of the United States, wearing a fur hat, inhabiting a planet with a moon, or inhabiting a world where nothing goes faster than light. Also, his size and shape and composition are accidental, not essential, to him. He could have been taller, he could have been slimmer, he could have had more or fewer fingers on his hands.
> Consider the last. He could have had six fingers on his left hand. There is a world that so represents him. We are supposing now that representation *de re* works by trans-world identity. So Humphrey, who is part of this world and here has five fingers on the left hand, is also part of some other world and there has six fingers on his left hand. *Qua* part of this world, he has five fingers, *qua* part of that world he has six.» (Lewis 1986b: 199)

O número de dedos da (única!) mão esquerda de Humphrey é um exemplo de um acidente intrínseco. No mundo atual, Humphrey tem apenas cinco dedos, mas poderia ter seis (ou mais). De acordo com a análise que estamos a considerar, dizer isto equivale a dizer que Humphrey no mundo atual tem cinco dedos na mão esquerda, mas num outro mundo tem seis na mesma mão. Como o indivíduo que é idêntico a Humphrey no mundo atual é exatamente o mesmo que o indivíduo que é idêntico a Humphrey no outro mundo possível, então Humphrey tem apenas cinco dedos na mão esquerda e tem, ao mesmo tempo, seis dedos nessa mão. Chegamos assim a um absurdo.

Uma resposta a este problema passa pela ideia de que a mão de Humphrey não tem cinco dedos absolutamente. Em vez disso, tem cinco dedos *no mundo atual* e tem seis dedos *num outro mundo possível*. É preciso, assim, aplicar-se um modificador à

atribuição de propriedades a um objeto, pelo menos quando há variabilidade ao longo dos mundos. Três formas de compreender esses modificadores são consideradas por D. Lewis. (1) Por vezes, os modificadores levam-nos a considerar as qualidades intrínsecas de uma parte, em vez das do todo: «If a tower is square on the third floor and round on the fourth floor, no worries; it's just that one segment differs in cross-sectional shape from another.» (Lewis 1986b: 200) Interpretando assim os modificadores usados antes, diríamos então que uma parte de Humphrey que está no mundo atual tem uma mão esquerda com cinco dedos, e uma outra parte de Humphrey que está noutro mundo possível tem uma mão esquerda com seis dedos. Esta hipótese, no entanto, foge à questão crucial: o que está em causa é a hipótese de Humphrey estar integralmente, e não parcialmente, em mais do que um mundo. (2) Uma coisa pode ter certas propriedades de acordo com uma fonte de representação, e não de acordo com outra: «If a man is honest according to the *News*, and crooked according to the *Times*, no worries; different papers tell different stories about him, they represent him differently and, at least one of them gets it wrong.» (Lewis 1986b: 200) Esta é a maneira como um proponente do realismo modal ersatz, de acordo com o qual os mundos possíveis são representações do mundo concreto, pode encarar os modificadores que estamos a considerar. Diferentes mundos possíveis ersatz podem representar um mesmo indivíduo – Humphrey, por exemplo – como tendo uma natureza intrínseca diferente. Mas neste caso Humphrey não tem efetivamente uma natureza intrínseca diferente em vários lugares a que chamamos 'mundos possíveis'. Esta alternativa não está disponível para quem defende a existência de uma pluralidade de mundos genuínos. Temos ainda que (3) um objeto pode ter propriedades extrínsecas contrárias relativamente a diferentes objetos: «If a man is father of Ed and son of Fred, no worries; he bears different relations to different individuals, and the extrinsic properties he thereby has – being a father, being a son – are compatible. Likewise if the wisest man in the village is by no means the wisest man in the nation». (Lewis 1986b: 200) Conceber esta possibilidade como a interpretação correta dos modificadores levaria a que se considerassem todos os acidentes como propriedades extrínsecas relacionais – esta categoria iria assim incluir o tamanho, a forma e a composição de um objeto. Ter uma mão com cinco dedos seria então uma relação que

Humphrey tem com o mundo atual, mas não com outro mundo com o qual estabelece a relação de ter uma mão com seis dedos. Para D. Lewis, esta ideia é inaceitável, e conclui então que não podemos ter identidade entre mundos – pelo menos se quisermos que essa identidade explique a modalidade *de re* (Lewis 1986b: 200-01).

Uma outra maneira de interpretar (3*) e (4*) passa por dizer que um objeto que existe apenas num mundo pode satisfazer *in absentia* frases abertas noutros mundos. Um objeto pode, por exemplo, satisfazer vicariamente uma frase aberta num mundo através da existência nesse mundo de um outro objeto que de alguma maneira o representa e que satisfaz efetivamente essa frase aberta. A hipótese que D. Lewis aceita é exatamente esta. Mais concretamente, considera que, estritamente falando, um indivíduo existe integralmente apenas num mundo, mas pode ter *contrapartes* noutros (Lewis 1968, 1973b: 39-40, 1986b: 8-10). É em virtude das suas contrapartes espalhadas por vários mundos que podemos dizer que um indivíduo existe em mais do que um mundo (num sentido menos estrito), e que tem diferentes características em cada um deles. Imaginemos um cão atual chamado Fido. É natural dizermos que Fido é essencialmente um cão, e que por isso satisfaz (3) e (3*). Como é que isso acontece? De acordo com a teoria das contrapartes, Fido existe apenas no mundo atual, mas existem diferentes cães noutros mundos que são contrapartes de Fido, e representam-no nesses mundos. (É conveniente até dizer que o nome ‘Fido’ denota esses cães nos respetivos mundos – é assim que podemos dizer que ‘Fido’ é alguma coisa como um designador rígido, apesar de, estritamente falando, não existirem tais designadores na teoria de D. Lewis, pelo menos para indivíduos particulares.) Essas contrapartes são, sem exceção, caninas, e é por esse motivo que Fido satisfaz (vicariamente e *in absentia*) a frase aberta ‘*x* é um cão’ em todos os mundos em que tem contrapartes e, consequentemente, é essencialmente um cão. Fido tem como cor o branco, mas poderia ter tido o cinzento e, deste modo, satisfaz (4) e (4*). Acontece o mesmo neste caso, *mutatis mutandis*: Fido tem contrapartes noutros mundos que são cinzentos, e é assim que ele satisfaz noutros mundos a frase aberta ‘*x* é cinzento’ e, consequentemente, poderia ter sido cinzento. Chegamos então a uma análise mais profunda de (3) e (4) do que (3*) e (4*):

(3**) Em todos os mundos em que há contrapartes de *x*, todas as contrapartes de *x* são um cão.

(4**) Em alguns mundos existem contrapartes de *x* que são cinzentas.

Esta teoria da modalidade *de re* afirma então que uma característica é essencial a uma coisa se e só se for partilhada por essa coisa e todas as suas contrapartes espalhadas pelo espaço lógico. Afirma também que uma característica é acidental a uma coisa se e só se for exemplificada por algumas das suas contrapartes, mas não por todas. É possível ainda definir a *essência* de uma coisa como a interseção das suas características essenciais: aquela característica que é exemplificada apenas por uma coisa e todas as suas contrapartes.

É assim que D. Lewis explica a relação de contraparte, em "Counterpart Theory and Quantified Modal Logic" (1968):

> «The counterpart relation is our substitute for identity between things in different worlds. Where some would say that you are in several worlds, in which you have somewhat different properties and somewhat different things happen to you, I prefer to say that you are in the actual world and no other, but you have counterparts in several other worlds. Your counterparts resemble you closely in content and context in important respects. They resemble you more closely than do the other things in their worlds. But they are not really you. For each of them is in his own world, and only you are here in the actual world.» (Lewis 1968: 114)

Assim como é entendida por D. Lewis, a relação de contraparte é uma relação de semelhança em aspetos relevantes: *y* é a contraparte de *x* no mundo *w* se *x* for o habitante de *w* que mais se assemelha (em aspetos relevantes) a *x*.[15] Os aspetos relevantes podem incluir tanto o conteúdo como o contexto de uma coisa – i. e., há que ter em conta tanto alguns aspetos da natureza intrínseca de uma coisa, a maneira como ela é em si mesma, assim como algumas características relacionais, por exemplo as suas origens e o seu papel histórico. Há que estabelecer por isso uma diferença entre os duplicados de um objeto em vários mundos, por um lado, e as suas contrapartes, por outro. Dois objetos são duplicados

15 Esta afirmação é uma generalização, mas não deve ser entendida como uma regra. As seguintes situações são ainda admitidas: (1) uma coisa pode ter mais do que uma contraparte num outro mundo, se duas coisas lhe forem igualmente semelhantes, e (2) uma coisa pode não ter nenhuma contraparte num certo mundo, se nenhuma coisa se parecer suficientemente com ela (Lewis 1968: 116).

quando partilham exatamente as mesmas qualidades intrínsecas, ainda que estabeleçam relações muito diferentes com o contexto em que se encontram. No caso de identificarmos as contrapartes com os duplicados, deixaríamos de dizer, por exemplo, que cada um de nós poderia ter sido muito diferente nas suas características anatómicas, o que à primeira vista é absurdo. Qualquer um de nós podia ter sido mais gordo do que é, ou podia ter ficado sem um dos membros, ou podia ser constituído por diferentes átomos, entre outras coisas. Em todo o caso, queremos que as propriedades extrínsecas sejam relevantes para determinar o que conta ou não como uma contraparte. Em geral, queremos dizer que coisas com exatamente as mesmas origens são contrapartes, ainda que difiram muito noutros aspetos ao longo da sua existência, e queremos ao mesmo tempo dizer que coisas muito parecidas entre si a nível intrínseco não têm de ser obrigatoriamente consideradas contrapartes, tendo em conta algumas diferenças que possam ter a nível extrínseco.

Nesta teoria, o propósito da relação de contraparte é simular a relação de identidade entre coisas em vários mundos. Ainda assim, o paralelismo entre ambas as relações é limitado, tendo em conta as diferenças entre elas quanto às características formais de cada uma. Mais concretamente, a relação de contraparte apresenta uma flexibilidade que não encontramos na relação de identidade. Por exemplo, a identidade é ao mesmo tempo uma relação transitiva, simétrica e reflexiva.[16] A relação de contraparte, por sua vez, é reflexiva (uma coisa é uma das contrapartes dela mesma), mas não é nem simétrica nem transitiva. Imaginemos que *b* é o objeto no mundo w_1 que mais se assemelha a *a* em @ (o mundo atual), que *c* é o objeto em w_2 que mais se assemelha a *b*, mas não é o objeto em w_2 que mais se assemelha a *a*. Deste modo, *c* é uma contraparte de *b*, que por sua vez é uma contraparte de *a*, sem que *c* seja uma contraparte de *a*. A relação não é, por isso, transitiva (Lewis 1968: 115). Imaginemos agora que um objeto *c* em w_1 é aquele que nesse mundo mais assemelha tanto a *a* como a *b*, ambos em @. Apesar disso, *c* assemelha-se mais a *b* do que a *a*, ao ponto de fazer com que *b*, mas não *a*, seja uma contraparte de *c*. Nesse caso, *c* é uma contraparte de *a*, mas *a* não é uma contraparte de *c*, e por isso mesmo a relação não é simétrica (Lewis 1968: 116). As características da

16 Uma relação R é transitiva se, para todo o *x*, *y* e *z*, R(*x*, *y*) e R(*y*, *z*), então R(*x*, *z*). Uma relação R é simétrica se, para todo o *x* e *y*, R(*x*, *y*), então R(*y*, *x*). E uma relação R é reflexiva se para todo o *x*, R(*x*, *x*). Todas estas são, obviamente, características da relação de identidade.

relação de contraparte permitem ainda algumas outras situações que não são permitidas pela relação estrita de identidade. (1) Duas coisas que habitam no mesmo mundo podem ter uma contraparte em comum num outro mundo. Uma pessoa meramente possível pode assemelhar-se igualmente a dois gémeos atuais. Deste modo, admitindo que essa pessoa é uma contraparte de um dos gémeos, então essa pessoa é uma contraparte de ambos. (2) Uma coisa pode ter duas contrapartes diferentes num mesmo mundo. Invertendo o exemplo anterior, podemos imaginar dois gémeos num mundo meramente possível que se assemelham igualmente a uma pessoa atual. Se um dos gémeos é uma contraparte dessa pessoa, então ambos são, e a pessoa tem duas contrapartes no mesmo mundo.

Uma das objeções mais comuns à teoria das contrapartes é que se não houver identidade entre coisas em vários mundos, quando dizemos que uma coisa pode ser desta ou daquela maneira, estamos a falar, de facto, daquilo que acontece a uma coisa completamente diferente numa situação contrafactual, o que parece – pelo menos à primeira vista – não ter qualquer relevância para aquilo que é ou não possível acerca de uma coisa no mundo atual. Consideremos, por exemplo, esta passagem de Saul Kripke em *Naming and Necessity* (1980):

> «The counterpart of something in another possible world is *never* identical with the thing itself. Thus if we say 'Humphrey might have won the election (if only he had done such-and-such),' we are not talking about something that might have happened to *Humphrey* but to someone else, a "counterpart". Probably, however, Humphrey could not care less whether someone else, no matter how much resembling him, would have been victorious in another possible world.» (Kripke 1980: 45)

A resposta de D. Lewis é que a teoria das contrapartes não nega de maneira nenhuma que é a Humphrey, ele mesmo e não uma outra coisa qualquer, que deve ser atribuída a possibilidade de ter vencido as eleições. O que esta teoria faz, em vez disso, é explicar como é que Humphrey tem essa propriedade, apelando à existência de pessoas meramente possíveis, não idênticas a Humphrey, que venceram efetivamente umas eleições nos mundos em que habitam e que se parecem suficientemente com ele para que a vitória delas seja relevante para que Humphrey tenha tido a possibilidade de vencer as eleições (Lewis 1986b: 196). O realismo modal com coisas que fazem parte de vários

mundos permite uma teoria alternativa para explicar em que consiste uma atribuição de propriedades modais, que está de acordo com a ideia intuitiva de que é por aquilo que acontece a uma coisa noutros mundos que ela tem certas propriedades essenciais e acidentais. Mas, como nota D. Lewis, mais nenhuma outra teoria permite dizer alguma coisa deste género, e neste aspeto não há qualquer vantagem entre o realismo modal com teoria das contrapartes e o ersatzismo:

> «I think counterpart theorists and ersatzers are in perfect agreement that there are other worlds (genuine or ersatz) *according to* which Humphrey – he himself! (stamp the foot, bang the table) – wins the election. And we are in equal agreement that Humphrey – he himself is not *part* of these worlds. Somehow, perhaps by suitable constituents or perhaps by magic, but anyhow not by containing Humphrey himself, the other world represents him as winning. […]
> Counterpart theory does say (and ersatzism does not) that someone else – the victorious counterpart – enters into the story of how it is that another world represents Humphrey as winning, and thereby enters into the story of how it is that Humphrey might have won. Insofar as the intuitive complaint is that someone else gets into the act, the point is rightly taken. But I do not see why it is any objection, anymore than it would be an objection against ersatzism that some abstract whatnot gets into the act. What matters is that the someone else, or the abstract whatnot, should not crowd out Humphrey himself. And there all is well.» (Lewis 1986b: 196)

Nem o realismo modal com contrapartes nem o ersatzismo afirmam que as coisas fazem parte de vários mundos (genuínos ou ersatz). Ambas as teorias, no entanto, admitem que uma coisa existe e tem propriedades diferentes *de acordo com* vários mundos. Na teoria das contrapartes, uma coisa existe *in absentia* em alguns mundos por ter neles uma ou mais do que uma contraparte. Assim como os mundos ersatz representam o mundo e as várias coisas como sendo de uma certa maneira, o mesmo acontece com os mundos genuínos aceites pelo realismo modal – com a diferença de, neste caso, a representação ser feita através da existência de coisas concretas.

1.5.3 – Essencialismo

Um enunciado que afirma que necessariamente ou possivelmente uma coisa tem estas ou aquelas características pode receber uma leitura *de dicto* ou *de re*. Quando

alguém afirma que necessariamente a pedra verde que está na mesa é uma pedra pode querer dizer (1) que é necessariamente uma verdade a afirmação (logicamente verdadeira, de facto) de que a pedra verde que está na mesa é uma pedra (leitura *de dicto*) ou (2) que a pedra verde que está na mesa é essencialmente uma pedra (leitura *de re*). O enunciado é verdadeiro em qualquer uma das interpretações, mas a ideia que é expressa de acordo com cada uma delas é diferente. A ideia *de dicto* é que aquilo que em cada mundo é a pedra verde que está na mesa é uma pedra, e a ideia *de re* é que exatamente a pedra verde que está na mesa no mundo atual é em qualquer mundo uma pedra. Na interpretação *de dicto* torna-se indispensável considerar a maneira como um objeto é descrito, mas o mesmo não acontece na interpretação *de re*. Imaginemos que a pedra verde que está em cima da mesa é numericamente idêntica à coisa em que Bill está a pensar. Esta identidade garante que a coisa em que Bill está a pensar é essencialmente uma pedra, mas não garante ainda assim que é necessariamente uma verdade a afirmação de que a coisa em que Bill está a pensar é uma pedra – de facto, podemos imaginar que a coisa em que Bill está a pensar é, por exemplo, uma árvore, uma coisa que certamente não é essencialmente uma pedra.

Admitir que a interpretação *de re* de certos enunciados modais é aceitável parece ser o mesmo que admitir o essencialismo: a teoria de acordo com a qual existe uma diferença entre as características essenciais e acidentais das coisas que é estabelecida independentemente da maneira como as descrevemos. A teoria das contrapartes, ao explicar as atribuições de propriedades essenciais e acidentais às coisas através das relações de semelhança que estabelecem com outras coisas possíveis, e encontrando assim uma análise para a modalidade *de re*, parece ser uma das variedades possíveis de essencialismo. Ainda assim, é apenas correto dizer que a teoria das contrapartes de D. Lewis é, no máximo, uma teoria essencialista relutante. Este é um dos comentários que D. Lewis faz num *postscript* a "Counterpart Theory and Quantified Modal Logic" (1968):

> «I am by no means offering a wholehearted defense of "Aristotelian essentialism." For the essences of things are settled only to the extent that the counterpart relation is, and the counterpart relation is not very settled at all. Like any relation of comparative overall similarity, it is subject to a great deal of indeterminacy (1) as to which respects of similarity and difference are to count at all, (2) as to the

> relative weights of the respects that do count, (3) as to the minimum standard of similarity that is required, and (4) as to which we eliminate candidates that are similar enough when they are beaten by competitors with stronger claims.
> Further, as with vagueness generally, the vagueness of the counterpart relation – and hence of essences and *de re* modality generally – may be subject to pragmatic pressures, and differently resolved in different contexts. The upshot is that it is hard to say anything false about essences. For any halfway reasonable statement will tend to create a context that (partially) resolves the vagueness of the counterpart relation in such a way as to make that statement true in that context. So almost anything goes. The true-hearted essentialist might well think me a false friend, a Quinean sceptic in essentialist's clothing.
> To take one extreme, a suitable context might deliver an antiessentialist counterpart relation – one on which anything is a counterpart of anything, and nothing has any essence worth mentioning. Or […] we might somehow partition things into kinds, and take a counterpart relation on which anything is a counterpart of anything of its kind. That would make the essence of a thing simply be its kind. […]
> At the opposite extreme, a suitable context might deliver a hyperessentialist counterpart relation – one on which nothing has any counterpart except itself. […] This counterpart relation, of course, is simply identity.» (Lewis 1983c: 42-3)

Ao explicar a modalidade *de re* em termos de uma relação de semelhança entre coisas em vários mundos, a teoria das contrapartes está a trabalhar com uma base arenosa e movediça. Apesar de esta teoria fornecer uma receita geral clara para a análise das atribuições de propriedades essenciais e acidentais, cumpre esta tarefa utilizando conceitos vagos, que vão deixar assim em aberto qual é a análise correta, de entre uma enorme variedade de análises possíveis, para cada caso particular. Ao dizermos que uma coisa tem essencialmente uma propriedade quando todas as coisas que com ela estabelecem uma relação adequada de semelhança – a relação de contraparte – também têm essa propriedade, deixamos ainda indeterminado qual é a relação adequada que deve ser adotada como a correta para interpretar cada afirmação modal *de re*. Não há apenas uma relação de contraparte, mas uma variedade delas, algumas mais inclusivas e outras mais restritas, que têm em consideração – e valorizam de uma maneira diferente – vários aspetos de semelhança e diferença entre as coisas. As possibilidades de escolha vão de um extremo até ao outro. De acordo com uma das relações de contraparte, todas as coisas são contrapartes umas das outras. É esta relação que estamos a utilizar quando dizemos de uma maneira aceitável que é acidental a um cão ser um cão e que poderia, por exemplo,

ter sido uma molécula de metano. De acordo com uma outra, nada é uma contraparte de uma coisa a não ser ela mesma. Utilizamos já uma relação muito próxima desta ao dizermos que um cão tem de ter exatamente o tamanho, a cor e a distância à lua que tem atualmente. Entre estes dois extremos, claro, há relações de contraparte mais convencionais: uma que atribui maior importância a fatores como a origem, outra à natureza intrínseca, outra ainda apenas ao que há de comum àquilo que pertence a um certo tipo natural, entre muitas outras alternativas. A escolha de uma destas relações de contraparte é feita em cada contexto, quando é feita, através de inúmeras influências pragmáticas – contando aqui fatores como as atitudes dos sujeitos que estão a atribuir essências às coisas. Em particular, um dos fatores que afeta a escolha de uma das várias relações de contraparte é a maneira como descrevemos as coisas (Lewis 1986b: 248-59). Esta variação permite acomodar certos casos de identidade contingente, como entre uma coisa e a matéria que a compõe. Por exemplo, podemos considerar que uma caixa e a porção de madeira de que é feita são atualmente idênticas (talvez porque queremos identificar as coisas que têm exatamente as mesmas partes e ocupam a mesma região espáciotemporal) e mesmo assim poderiam ter sido duas coisas numericamente diferentes, tendo em conta que os nomes comuns 'caixa' e 'porção de madeira' selecionam diferentes relações de contrapartes e, num mundo, existe uma coisa que é uma contraparte da caixa e outra que é uma contraparte da porção de madeira – e não são a mesma coisa. Outro exemplo, explorado em "Counterparts of Persons and Their Bodies" (1971), é o da relação entre uma pessoa e o seu corpo. Esta flexibilidade oferecida pela teoria das contrapartes permite a um materialista dizer que uma pessoa é idêntica ao seu corpo sem contrariar a ideia intuitiva de que cada pessoa poderia ter tido outro corpo. (Podemos encontrar uma abordagem semelhante a esta em (Gibbard 1975).)

É por isso que a variedade da teoria das contrapartes defendida por D. Lewis, mesmo admitindo como aceitável a noção de essência, não está comprometida com o essencialismo da mesma forma que está uma teoria que explica a modalidade *de re* através da identidade entre coisas em vários mundos. É aceitável dizer que uma relação de contraparte é a adequada para interpretar certas afirmações dependendo dos nossos interesses e atitudes, como também é aceitável dizer o mesmo da relação de semelhança

em geral, mas não da relação de identidade. Não há, pelo menos aparentemente, qualquer indeterminação naquilo que conta como idêntico: cada coisa é idêntica a si mesma e a nada mais. Se a essência de uma coisa é determinada pelas características que *essa mesma* coisa tem de uma maneira estável nos vários mundos, então a essência é alguma coisa determinada, que não é afetada por fatores pragmáticos. A flexibilidade da relação de contraparte comparativamente à de identidade é, de acordo com D. Lewis, uma das vantagens da teoria que propõe.

1.5.4 – Semântica

Um dos propósitos mais relevantes do uso de uma linguagem é transmitir e receber informação. De acordo com D. Lewis, isso acontece deste modo:

> «Suppose (1) that you do not know whether *A* or *B* or …; and (2) that I do know; and (3) that I want you to know; and (4) that no extraneous reasons much constrain my choice of words; and (5) that we both know that the conditions (1)-(5) obtain. Then I will be truthful and you will be trusting and thereby you will come to share my knowledge, I will find something to say that depends for its truth on whether *A* or *B* or … and that I take to be true. I will say it and you will hear it. You, trusting me to be willing and able to tell the truth, will then be in a position to infer whether *A* or *B* or ….» (Lewis 1980a: 80)

Vamos supor que um falante pretende transmitir a um ouvinte a informação de que a neve é branca. Este processo vai exigir que o falante escolha emitir uma expressão que saiba que é verdadeira se e só se a neve é branca. Em certas circunstâncias, o ouvinte vai confiar naquilo que foi dito pelo falante, e por isso se ele tiver conhecimento de que a expressão emitida é verdadeira se e só se a neve é branca, vai passar a saber que a neve é branca, recebendo assim a informação que o falante pretendia transmitir. A verdade das expressões utilizadas não depende, no entanto, apenas de questões de facto, mas também da interpretação que lhes é atribuída – e esta varia de comunidade para comunidade. Em português, 'A neve é branca' serve para expressar aquilo que o falante queria dizer. Mas pensemos no exemplo, apresentado em (Lewis 1980a: 80), da tribo dos mentirosos. Os membros desta tribo dizem apenas aquilo em que acreditam e usam para isso uma linguagem parecida com o português, mas em que os valores de verdade das frases estão

invertidos. Para serem sinceros, os mentirosos devem dizer 'A neve não é branca' se querem expressar a ideia de que a neve é branca. Acontece, assim, que a comunidade de falantes do português e a tribo dos mentirosos atribuem diferentes interpretações às expressões que usam. D. Lewis propõe (1969: cap. 5, 1975: 7-12, 1986b: 40) que a interpretação correta das expressões da linguagem que é usada numa comunidade é aquela que permite dizer que os membros dessa comunidade participam numa convenção de sinceridade e confiança (*ver* secção 3.5). Repare-se que existem várias convenções deste género: uma para cada interpretação diferente das expressões que podem ser usadas por uma comunidade. Tanto os falantes de português como os membros da tribo dos mentirosos participam numa convenção de sinceridade e confiança, mas não na mesma. A convenção mantida pelos falantes de português envolve usar a frase 'A neve é branca' para dizer que a neve é branca, enquanto que na convenção dos mentirosos, essa frase serve para dizer exatamente o contrário.

Existem várias convenções de sinceridade e confiança prevalecentes em várias comunidades. Para se descrever de uma maneira completa a prática linguística dos membros dessas várias comunidades, é preciso especificar as condições de verdade que cada uma delas atribui às frases da linguagem usada pelos seus membros. D. Lewis considera que esta tarefa deve ser desempenhada por uma gramática – uma estrutura que abrange tanto aspetos sintáticos como semânticos da linguagem.

Uma gramática envolve (1) a atribuição de um valor semântico e de uma categoria a um conjunto finito de expressões básicas, (2) a especificação de regras sintáticas que nos dizem que combinações de expressões de certas categorias são permitidas e a que categoria pertencem as expressões complexas assim obtidas, e (3) a especificação de regras semânticas que determinam o valor semântico das expressões complexas a partir do valor dos seus constituintes imediatos e da maneira como estes estão combinados. (Uma das categorias de qualquer gramática é a das frases, básicas ou não.) Pode também envolver (4) um componente transformacional: uma série de procedimentos que atribuem a certas expressões obtidas a partir do léxico básico uma representação linguística – marcas de tinta, por exemplo. As expressões do léxico básico – e aquelas que são obtidas a partir destas através das regras sintáticas – de uma gramática com um componente

transformacional não têm de ser as palavras da linguagem que essa gramática está a descrever. (*Ver* anexo 4, onde é apresentado superficialmente um exemplo de gramática.)

Um valor semântico é qualquer coisa que, quando é atribuído por uma gramática a uma expressão, desempenha a tarefa de contribuir composicionalmente para a determinação do valor semântico das expressões complexas de que essa expressão é um constituinte imediato, juntamente com os valores semânticos dos restantes constituintes e das regras semânticas que lhes estão associadas. Em geral, o propósito de atribuir valores às expressões é chegar especificamente aos valores das frases, que se pretende adicionalmente que especifiquem as condições de verdade da frase a que estão associados (Lewis 1980a: 82-4, 1986b: 41). Note-se que 'valor semântico' é uma palavra neutra que D. Lewis usa para falar daquilo que pode ocupar este papel teórico, sem se comprometer à partida com o tipo de coisa que deve ocupá-lo – preferindo não falar, neste contexto, em coisas como referência, denotação, sentido, intensão, extensão, conceitos, expressão, representação, carácter, entre outros bastante comuns na literatura filosófica.

Mundos e indivíduos possíveis são úteis na construção dos valores semânticos. A verdade das frases depende de questões de facto contingentes – que variam de mundo para mundo. (Excetuam-se aqui as frases necessárias e impossíveis, respetivamente verdadeiras e falsas em qualquer mundo.) Daqui surge a hipótese de conceber as funções de mundos para valores de verdade (0 e 1, por exemplo) como os valores semânticos adequados às frases. Essas funções codificam a informação sobre a verdade das frases nos vários mundos: cada função vai atribuir o valor 1 aos mundos em que a frase a que está atribuída é verdadeira e o valor 0 aos restantes. Ao especificarmos cada uma dessas funções, estamos automaticamente a especificar as condições de verdade da frase correspondente: dizemos que a função atribui o valor 1 aos mundos com tais e tais características, o que implica que a frase é verdadeira se e só se o mundo atual tem tais e tais características.

Esta hipótese é, ainda assim, incompleta. A verdade de algumas frases, chamadas *indexicais*, não depende apenas de como o mundo é, mas também de como é o contexto em que a frase é usada. 'Esta estrada vai para Espanha' é verdadeira no mundo atual se, por exemplo, um falante apontar num mapa para uma estrada que efetivamente vai para

Espanha, mas é falsa neste mesmo mundo se o falante apontar para uma estrada na Colômbia. O que varia entre estes dois casos não é o mundo em que o falante se encontra, mas o ponto de vista que o falante tem do mundo em que habita. As funções de mundos para valores de verdade são insensíveis à variação da verdade de uma frase em contextos diversos que se encontram no interior de um mesmo mundo, e por isso não servem para codificar toda a informação relevante acerca das condições de verdade de todas as frases, mas apenas das *eternas*.

Assumindo uma pluralidade de mundos, as variações da verdade entre mundos e entre contextos não têm de ser concebidas como dois fatores distintos, como explica D. Lewis:

> «When truth-in-English depends on matters of fact, that is called *contingency*. When it depends on features of context, that is called *indexicality*. But need we distinguish? Some contingent facts are facts about context, but are there any that are not? Every context is located not only in physical space but also in logical space. It is at some particular possible world – our world if it is an actual context, another world if it is a merely possible context. […] It is a feature of any context, actual or otherwise, that its world is one where matters of contingent fact are a certain way. Just as truth-in-English may depend on the time of the context, or the speaker, the standards of precision, or the salience relations, so likewise may it depend on the world of the context. Contingency is a kind of indexicality.» (Lewis 1980a: 82)

Há uma enorme variedade de características semanticamente relevantes de um contexto. Em (Lewis 1970b: 24) são listadas as seguintes: (1) mundo, (2) tempo, (3) lugar, (4) falante, (5) audiência, (6) objetos indicados e (7) discurso prévio. Aí, D. Lewis propõe que o valor semântico (a expressão usada nessa altura era 'intensão') de uma frase seja uma função de *índices* para valores de verdade. Um índice é uma sequência de características de um contexto – chamadas *coordenadas* do índice. Admitindo (1)-(7) como um inventário completo das características relevantes, um índice é identificado com um séptuplo ordenado que contém um mundo, um instante de tempo, um lugar, um falante, uma audiência, um objeto indicado e discurso prévio. Mais à frente no mesmo artigo (1970b: 63-4), considera ainda expandir os índices para incluírem mais duas coordenadas.

(8) Objetos proeminentes num contexto. Quando dizemos 'A porta está aberta' não temos de estar a falar, como nota D. Lewis, da única porta existente no mundo atual e no momento presente, nem de uma porta perto do lugar em que está a falar, nem de uma porta para a qual se está a apontar, nem de uma porta mencionada anteriormente. A ideia é que há uma porta que é de alguma maneira a mais saliente naquele contexto, mas ela pode adquirir esse estatuto de maneiras variadas e até complexas que podem envolver aspetos mentais do falante e da audiência.

(9) História causal da aquisição dos nomes. Um nome não tem de adquirir um referente por alguma descrição desse objeto que o falante tem em mente e que associa ao nome. Em vez disso, pode adquirir um referente através de uma cadeia causal que se estende desde o momento em que o objeto foi batizado até que o nome tenha sido adquirido pelo falante. Por isso, diferentes histórias causais que desencadeiam a aquisição de um nome podem associá-lo a diferentes objetos. Este aspeto pode ser encarado como mais um fator contextual.

Em "Index, Context and Content" (1980a), no entanto, D. Lewis desiste da estratégia de construir índices com as coordenadas suficientes para capturar as condições de verdade das frases de uma linguagem, por ver essa tarefa como extremamente complexa. Em vez disso, propõe que se substituam os índices por *contextos*. Um índice é apenas um aglomerado de características de qualquer contexto atual ou possível. Um contexto, por sua vez, é uma localização específica no espaço lógico em que uma frase pode ser emitida. Existem índices com características que nenhum contexto tem. Por exemplo, nenhum contexto pode estar localizado num certo mundo e envolver um falante que não existe nesse mundo – esta ideia é absurda –, mas um índice pode conter ao mesmo tempo um falante e um mundo em que ele não existe. Além disso, um índice carrega consigo a informação oferecida pelas várias coordenadas e nada mais. Um contexto, por sua vez, pode carregar uma enorme quantidade de informação implícita. Em (1980a: 85), D. Lewis identifica cada contexto com um triplo ordenado de um mundo, um tempo e um lugar, mas acrescenta que existem inúmeras características de um contexto para além destas que ficam explicitamente especificadas, todas elas determinadas pelo carácter intrínseco ou relacional da localização oferecida por estas três coordenadas. Por exemplo,

o falante de cada contexto é o falante que se encontra no lugar, momento e mundo especificados no contexto, a audiência de um contexto é o conjunto de pessoas com quem o falante do contexto está a comunicar, e por aí em diante. A cada contexto corresponde um índice que contém as características desse contexto. Mas uma diferença crucial entre um contexto e o índice que lhe corresponde é que se alterarmos uma das características do índice obtemos um novo índice diferente, mas se alterarmos uma das características do contexto – para além do mundo, tempo e lugar – não obtemos um novo contexto.

Mais tarde, em (Lewis 1983c: 230), é proposta a ideia peculiar, ainda que simples, de identificar um contexto com um segmento momentâneo de um falante possível. (Esta proposta depende para a sua adequação de duas teses controversas aceites por D. Lewis: (1) um indivíduo possível existe apenas num dos mundos, e (2) existem partes temporais de um objeto que perdura ao longo do tempo.) Neste caso, o mundo do contexto é o mundo em que o falante se encontra, o tempo do contexto é o instante em que esse falante existe, e por aí em diante.

Funções de contextos para valores de verdade servem para capturar as condições de verdade das frases. Não servem, ainda assim, como valores semânticos apropriados às frases. Recordemos que o valor semântico de uma expressão tem de carregar a informação necessária para determinar composicionalmente os valores semânticos das expressões complexas de que essa expressão faz parte. Para cumprir essa tarefa, é preciso que os valores semânticos integrem também os índices. Para ver porquê, vamos considerar uma gramática extremamente simples, e assumir que o valor semântico que essa gramática atribui a cada uma das suas frases é uma função de contextos para valores de verdade. O nosso exemplo de gramática vai conter expressões de apenas três categorias: frases, modificadores e conectivos. Os modificadores e os conectivos permitem-nos obter novas frases a partir de outras mais básicas. A única diferença entre eles é que os modificadores formam uma frase a partir de uma outra, enquanto que os conectivos formam uma frase ao ligar outras duas (ou mais, mas não há problema em ignorar aqui essa possibilidade). De maneira a cumprir a sua tarefa composicional de uma forma simples, os modificadores têm como valor semântico apropriado uma função de valores semânticos de uma frase para valores semânticos de uma frase, e aos conectivos é atribuída uma função de pares

ordenados de valores semânticos de frase para valores semânticos de frase. (*Ex hypothesi*, estamos a falar de funções que atribuem funções de contextos para valores de verdade, ou a pares de ordenados destas, outras funções de contextos para valores de verdade.) Especificamos agora os elementos desta gramática.

Vocabulário básico
(1) 'Os cães são mamíferos' (frase).
(2) 'Eu existo' (frase)
(3) 'Não é verdade que' (modificador)
(4) 'Necessariamente' (modificador)
(5) 'No passado é verdade que' (modificador)
(6) 'e' (conectivo)

Regra sintática para os modificadores: se ϕ é uma frase e η é um modificador, $\eta + \phi$ é uma frase.

Regra sintática para os conectivos: se ϕ e ψ são frases e η é um conectivo, $\phi + \eta + \psi$ é uma frase.

(As únicas expressões complexas que fazem parte do léxico desta gramática são frases.)

Exemplos de frases complexas obtidas a partir do vocabulário básico
(7) 'Não é verdade que os cães são mamíferos'
(8) 'Necessariamente os cães são mamíferos'
(9) 'No passado é verdade que os cães são mamíferos'
(10) 'Os cães são mamíferos e eu existo'
(11) 'Não é verdade que não é verdade que os cães são mamíferos'
(12) 'Não é verdade que necessariamente os cães são mamíferos'
...
(13) 'Não é verdade que eu existo'
(14) 'Necessariamente eu existo'
(15) 'No passado é verdade que eu existo'
...
(16) 'No passado é verdade que não é verdade que eu existo'
...

Regra semântica para os modificadores: se ϕ é uma frase com valor v e η é um modificador com valor v, o valor de $\eta + \phi$ é $u(v)$.

Regra semântica para os conectivos: se ϕ é uma frase com valor v, ψ uma frase com valor w e η um conectivo com valor u, o valor de $\phi + \eta + \psi$ é $u(v, w)$.

(Repare-se que os valores semânticos apropriados aos modificadores e aos conectivos foram concebidos para que estas regras semânticas simples fossem adequadas.)

Assumimos há pouco que o valor semântico apropriado às frases é uma função de contextos para valores de verdade. Com isto em mente, podemos já especificar facilmente os valores de (1) e (2). O valor semântico de (1) é uma função que atribui a cada contexto o valor de verdade 1 se e só se no mundo em que esse contexto se encontra os cães são mamíferos. A (2) associamos a função que atribui a cada contexto o valor de verdade 1 se só se o falante que faz parte desse contexto existe nesse contexto (com o método das contrapartes dizemos antes: se só se existir nesse contexto uma contraparte do falante desse contexto, mas essa modificação é irrelevante neste momento). Resta-nos agora especificar os valores semânticos de (3)-(6), para que desse modo fiquem implicitamente especificados o valor semântico de todas as expressões que fazem parte do léxico desta gramática – e consequentemente as condições de verdade de todas as frases. É relativamente fácil especificar valores semânticos adequados para (3) e (6). Estes são operadores verofuncionais, que formam frases cujo valor de verdade depende unicamente do valor de verdade dos constituintes imediatos. O valor semântico destes operadores tem apenas de determinar o valor de verdade que a frase resultante deve ter num contexto a partir do valor de verdade que as frases constituintes têm nesse contexto. A (3) atribuímos a função que associa qualquer função *f* de contextos para valores de verdade a uma outra função de contextos para valores de verdade que associa a um contexto o valor 1 se e só se *f* associa a esse contexto o valor 0. De uma maneira semelhante, a (6) atribuímos a função que associa pares de funções (*f, g*) de contextos para valores de verdade a uma função de contextos para valores de verdade que associa a um contexto o valor 1 se e só se tanto *f* como *g* associam a esse contexto o valor 1. Estes valores para (3) e (6) são adequados, e dão-nos as condições de verdade corretas de qualquer frase obtida através destes operadores, se as condições de verdade das frases constituintes estiverem também corretas. Assumindo que especificamos corretamente as condições de verdade de (1) e (2) temos implicitamente especificadas as condições de verdade corretas de frases como (7), (10), (11) e (13), entre muitas outras mais complexas.

Resta-nos ainda considerar os valores de (4) e (5). Vimos antes que um operador modal como (4) pode ser tratado como um quantificador aplicado a mundos possíveis. Conseguimos um valor semântico para (4) se tratarmos, em vez disso, um operador modal

como um quantificador sobre contextos possíveis. Temos, então, para (4) uma função que associa a cada função *f* de contextos para valores de verdade uma outra função de contextos para valores de verdade que associa a qualquer contexto o valor de verdade 1 se e só se *f* atribui a todos os contextos o valor de verdade 1. Este parece levar-nos a resultados adequados para (8) e (12), deixando como verdade que necessariamente os cães são mamíferos, e como uma falsidade a negação dessa ideia. Mas leva-nos a resultados inaceitáveis para (14). Estamos a admitir como parte da nossa hipótese que um contexto é um falante, ainda que momentâneo. Assim, em qualquer contexto existe o falante que corresponde a esse contexto, e por essa razão em qualquer contexto é verdade a frase 'Eu existo'. Com os valores que associamos a (2) e (4) isso implica que 'Necessariamente eu existo' é verdadeira em qualquer contexto – o que é absurdo e prova que há algo de errado nestes valores. O mesmo vai acontecer com (5). Uma hipótese é a de atribuir a (5) uma função que associa uma qualquer função *f* de contextos para valores de verdade uma outra função de contextos para valores de verdade que associa a um contexto *c* o valor 1 se e só se existe um contexto *d* que seja anterior a *c*, pertençam ambos ao mesmo mundo, e *f* atribui a *d* o valor 1. Esta é a melhor tentativa para encontrar um valor para (5) com os recursos disponíveis. Mas é uma tentativa falhada. Consideremos (16). Esta frase, de acordo com estes valores, é falsa em qualquer contexto – o que é mais uma vez absurdo. Como em todos os contextos há um falante, também é verdade que num contexto anterior a qualquer outro há um falante.

Estes resultados mostram a inadequação destes valores para lidar com a necessidade de considerar o valor de verdade que certas frases teriam caso algumas das características do contexto fossem alteradas, enquanto outras se mantinham fixas. É por isso que os índices devem ser integrados nos valores semânticos: em vez de funções de contextos para valores de verdade devemos atribuir funções de sequências formadas por um contexto e por algumas coordenadas – aquelas que forem necessárias – de um índice. (Também podíamos atribuir funções de pares ordenados de contextos e índices, com todas as coordenadas no seu interior, para valores de verdade. As duas alternativas são equivalentes e não há qualquer relevância em optar por uma delas.) Para lidar com os modificadores (4) e (5) precisamos apenas de uma coordenada modal e uma coordenada

temporal. Podemos então assumir que um valor semântico para uma frase é uma função de triplos ordenados formados por um contexto, um mundo e um tempo. Vamos reformular os valores de (1) e (2). O valor de (1) passa a ser uma função que associa a cada triplo (*c*, *w*, *t*) o valor de verdade 1 se e só se os cães são mamíferos em *t* no mundo *w*. Por sua vez, o valor de (2) passa a ser uma função que associa a cada triplo (*c*, *w*, *t*) o valor de verdade 1 se e só se o falante que se encontra em *c* existe em *t* no mundo *w*. A quantidade de informação contida neste novo valor semântico para (2) é muito maior que a que estava contida no anterior. Agora, este valor tem alguma coisa a dizer sobre a verdade da frase noutros mundos e noutros tempos, mantendo-se ainda assim fixo o falante do contexto que se está a considerar. Facilmente são ajustáveis os valores ds operadores verofuncionais (3) e (6). Quanto a (4) e (5) estamos agora em condições de apresentar valores adequados. Os operadores modais podem ser entendidos mesmo como quantificadores sobre mundos, e nesse sentido o valor de (4) é uma função que associa uma função *f* de triplos formados por um contexto, um mundo e um tempo a uma outra função do mesmo género que associa a qualquer triplo (*c*, *w*, *t*) o valor de verdade 1 se e só se, para todos os mundos *v*, *f* atribui a (*c*, *v*, *t*) o valor de verdade 1. Por fim, o valor de (5) é também uma função que associa uma função *f* de triplos formados por um contexto, um mundo e um tempo a uma outra função do mesmo género que associa a cada triplo (*c*, *w*, *t*) o valor de verdade 1 se e só se existe um tempo *u* que é anterior a *t* e *f* atribui a (*c*, *w*, *u*) o valor de verdade 1. Com estes valores, deixam de existir os problemas levantados acima.

1.5.5 – Propriedades, relações e proposições

Uma outra vantagem que D. Lewis encontra no realismo modal é a de tornar adequada uma teoria extremamente

e simples acerca das propriedades, relações e proposições, que identifica cada uma destas entidades com um certo conjunto (Lewis 1986b: 60-6). De acordo com esta teoria, por um lado, uma propriedade ou relação é o conjunto dos seus exemplares. As propriedades são, então, conjuntos de objetos singulares e as relações são conjuntos de sequências de

objetos.[17] (Relações binárias são conjuntos de pares ordenados, relações ternárias de triplos ordenados, e por aí em diante. D. Lewis admite ainda relações que têm como elementos sequências de comprimento variável.) E, também de acordo com esta teoria, por outro lado, uma proposição é o conjunto de mundos em que é verdadeira. Nesta conceção, as proposições passam a ser um tipo particular de propriedades, exemplificadas apenas por mundos inteiros, não havendo diferença, por exemplo, entre a proposição *que a neve é branca* e a propriedade *ser um mundo possível em que a neve é branca*. Esta ideia traz unidade à teoria, que pode então ser resumida assim: uma propriedade, uma relação ou uma proposição é o conjunto dos objetos ou das sequências de objetos que a exemplificam. Existe, nesta teoria, uma propriedade que é idêntica a uma relação e a uma proposição: a propriedade (relação e proposição) impossível, que não é exemplificada por nenhuma coisa possível (nem por nenhuma sequência de coisas possíveis, nem por nenhum mundo possível). Esta entidade que é ao mesmo tempo uma propriedade, uma relação e uma proposição é identificada, claro, com o conjunto vazio.[18]

Ao falar-se dos objetos meramente possíveis, e não apenas dos atuais, como ingredientes na construção das propriedades e relações, evita-se a identidade entre propriedades e relações acidentalmente coextensivas – i. e., com os mesmos exemplares no mundo atual, ainda que com exemplares diferentes noutros mundos. Numa ontologia afetada por questões de facto contingentes, como a maior parte daquelas que excluem o

---

17 Temos de falar nas relações como conjuntos de sequências de objetos relacionados, e não como conjuntos de conjuntos de objetos, tendo em conta que a ordem em que os objetos aparecem é relevante. Imaginemos que *a* ama *b*, mas *b* odeia *a*. A relação binária *ter amor por* tem como elemento o par ordenado (*a*, *b*), mas não (*b*, *a*), que pertenceria por sua vez à relação *ter ódio por*. Sequências de objetos distinguem-se dos conjuntos precisamente por codificarem uma ordem. Dois conjuntos são idênticos se e só se tiverem os mesmos elementos. Deste modo, {*a*, *b*} é idêntico a {*b*, *a*}. No caso de um par ordenado (mas o mesmo pode ser dito, *mutatis mutandis*, das restantes sequências) pretende-se que (*x*, *y*) seja idêntico a (*w*, *z*) se e só se *x* é idêntico a *w* e *y* a *z*. Nesse caso, não basta que os pares tenham os mesmos elementos para serem o mesmo: os elementos têm ainda de aparecer na mesma ordem.

18 De uma maneira ainda mais simples, podemos apresentar assim esta teoria: uma propriedade é um qualquer conjunto de objetos que não tem apenas sequências como membros, uma relação é um qualquer conjunto que ou não tem qualquer membro ou tem apenas sequências como membros, e uma proposição é um qualquer conjunto que ou não tem qualquer membro ou tem apenas mundos inteiros como membros. Repare-se que (1) o conjunto vazio é a propriedade, relação e proposição impossíveis, (2) a cada conjunto que contenha pelo menos um objeto singular, corresponde a idêntica propriedade de pertencer a esse conjunto (e o mesmo se aplica, *mutatis mutandis*, às proposições como propriedades de mundos inteiros), e (3) a qualquer conjunto de sequências corresponde a idêntica relação de fazer parte de uma das sequências desse conjunto juntamente com outros objetos numa certa ordem (Lewis 1983b: 344-51).

que é meramente possível, os conjuntos, por um lado, e as propriedades e relações, por outro, comportam-se de maneira diferente: conjuntos com os mesmos elementos são numericamente idênticos, mas propriedades e relações com os mesmos exemplares podem ser diferentes. Assumindo que apenas existe o que é atual, temos de admitir que o conjunto de criaturas com coração é idêntico ao conjunto de criaturas com rins, tendo em conta que, no mundo atual, todas e apenas as criaturas com coração têm rins. É absurdo, no entanto, dizer-se que as propriedades *ter um coração* e *ter rins* são idênticas. Logo, ou se abandona o tratamento das propriedades como conjuntos de exemplares, ou se abandona a ontologia atualista que nos serviu de hipótese. Podemos chegar a esta mesma ideia através do exemplo das propriedades sem exemplares no mundo atual, como estas: *ser um unicórnio*, *ser uma árvore falante*, *habitar um mundo em que a mecânica clássica é correta*, entre outras. O conjunto correspondente a todas estas propriedades, se nos restringirmos ao mundo atual, é exatamente o mesmo – nomeadamente, o conjunto vazio. Reparemos ainda que estas propriedades seriam, nesta hipótese, idênticas ainda a propriedades impossíveis, sem exemplares em qualquer um dos mundos, como estas: *ser um unicórnio e não ser um unicórnio*, *não existir*, *habitar um mundo em que 2 + 2 = 5*, entre outras.

Esta abordagem tem de enfrentar ainda outras complicações. Em primeiro lugar, a pertença de um objeto a um certo conjunto é um facto que não varia de mundo para mundo. Assim, se ter uma propriedade é concebido como pertencer a um conjunto, estamos obrigados a dizer que qualquer objeto tem todas as suas propriedades essencialmente: em todos os mundos, um objeto pertence exatamente aos mesmos conjuntos. Em vez de se identificar as propriedades com o conjunto dos seus exemplares, seria mais adequado identifica-las, por exemplo, com funções que atribuem a cada mundo um conjunto de coisas que existem nesse mundo, entendidas como os exemplares dessa propriedade nesse mundo.

A teoria das contrapartes permite resolver esta complicação. De acordo com esta teoria, um objeto tem a propriedade F essencialmente se só se nos vários mundos todas as suas contrapartes têm F. Um objeto e as suas contrapartes não são numericamente idênticos, apesar de por vezes ser útil falar-se como se fossem. Ora, não sendo idênticos,

um objeto e as suas contrapartes podem ser membros de conjuntos diferentes. É por um objeto ter contrapartes que pertencem a conjuntos a que ele próprio não pertence que esse objeto não tem todas as suas propriedades essencialmente.

Em segundo lugar, há que ter em conta que algumas propriedades são instanciadas de uma maneira relativa. Por exemplo, uma pessoa que começa a sofrer uma dor num certo momento tem a propriedade *sentir dor* relativamente a certos instantes de tempo, mas não em relação a outros. Um bolo que é branco na cobertura, mas castanho no seu recheio tem a propriedade *ser branco* relativamente à sua cobertura, mas não a tem relativamente ao seu recheio. A pergunta que se deve colocar agora é: será que esta pessoa e este bolo pertencem ou não aos conjuntos identificados com as propriedades *sentir dor* e *ser branco*, respetivamente? Um objeto pertence ou não a um conjunto de um modo absoluto. Por esse motivo, independentemente do que se queira dizer acerca da exemplificação relativa de uma propriedade, não há espaço para isso acontecer se exemplificar uma propriedade for concebido como pertencer a um conjunto. A moral desta história é que ou não há exemplificação relativa, ou as propriedades não são conjuntos de exemplares. Talvez as propriedades sejam coisas como funções de mundos, ou instantes de tempo, ou lugares, para conjuntos de objetos que existem nesses mundos, nesses instantes ou nesses lugares, entendidos como os exemplares dessas propriedades relativamente a estes parâmetros.

A resposta de D. Lewis é que não há exemplificação relativa. As propriedades que são instanciadas relativamente a uma coisa, mas não relativamente a outra, não são propriedades genuínas, mas relações (Lewis 1986b: 53). Retomando os casos que nos serviram de exemplo, seja *a* uma pessoa com dor no momento *t*, e seja *b* o bolo de cor variável. *Sentir dor* é uma relação que é exemplificada pelo par ordenado (*a*, *t*), assim como *ser branco* é uma relação exemplificada por (*b*, cobertura de *b*), e *ser castanho* por (*b*, recheio de *b*). Há, ainda assim, propriedades genuínas envolvidas nestes exemplos, nomeadamente aquelas que são exemplificadas pelas partes em vez do todo. O bolo é branco numa das suas partes, e essa parte tem absolutamente a propriedade de ser branca. Da mesma maneira, tendo em conta que D. Lewis acredita que um objeto perdura no tempo por ter várias partes momentâneas, pode dizer que algumas das partes temporais

da pessoa sentem dor absolutamente. Ainda assim, não se pode dizer o mesmo em todos os restantes casos de suposta exemplificação relativa. 23 tem a propriedade de ser membro relativamente ao conjunto dos números naturais, mas não relativamente ao conjunto de gafanhotos. Nenhuma parte de 23 tem a propriedade de ser membro absolutamente (isso é impossível porque estamos a falar de uma relação), e por isso este exemplo é diferente dos anteriores. Em todos estes exemplos, no entanto, podemos ainda encontrar outra propriedade genuína: cada um dos objetos tem a propriedade de estar relacionado desta ou daquela maneira com um outro objeto específico (e tem também a propriedade de estar relacionado desta ou daquela maneira com *algum* objeto). A pessoa *a* tem a propriedade de sentir dor em *t*, o bolo *b* tem a propriedade de ser branco na cobertura, assim como a propriedade de ser castanho no recheio e, ainda, 23 tem a propriedade de ser membro do conjunto dos números naturais.

Em terceiro lugar, finalmente, há que notar que apesar de a teoria de D. Lewis evitar a inaceitável identificação de propriedades acidentalmente coextensivas, não consegue distinguir, ainda assim, as propriedades que são coextensivas de uma maneira necessária. Como vimos antes, conjuntos com os mesmos elementos são idênticos. Ainda que a ontologia do realismo modal seja suficientemente vasta para fazer com que propriedades diferentes não tenham exatamente os mesmos exemplares (no espaço lógico inteiro) por causa de questões de facto contingentes, não é capaz de evitar que isso aconteça em virtude de factos necessários. No caso de exatamente os mesmos objetos possíveis exemplificarem certas propriedades, estas têm de ser tratadas como idênticas. Tendo em conta que, necessariamente, todos e apenas os triângulos têm três ângulos, temos de dizer que as propriedades *ter três lados* e *ter três ângulos* são idênticas. Além disso, há apenas uma propriedade universalmente exemplificada por todos os possíveis e uma propriedade impossível: a primeira é o conjunto de todos os objetos possíveis, a última é o conjunto vazio. O mesmo se aplica às relações e, claro, às proposições, como um caso particular de propriedades. Proposições verdadeiras exatamente nos mesmos mundos são idênticas. E, em particular, há apenas uma proposição impossível e uma proposição necessária: a primeira é o conjunto vazio, e a última é o conjunto de todos os mundos. Tendo em conta que, necessariamente, alguns perus voam se e só se nem todos

os perus não voam, temos de admitir que as proposições *que alguns perus voam* e *que nem todos os perus não voam* são a mesma. Acontece a mesma coisa com qualquer par de proposições logicamente equivalentes – temos que qualquer proposição é idêntica à negação da sua negação, a conjunção de duas proposições P e Q é a mesma proposição que a negação da disjunção das negações de P e Q, e por aí em diante. Além disso, todos os teoremas em matemática e lógica, por exemplo, expressam a mesma proposição, assim como as verdades analíticas e (presumivelmente) as afirmações que fazem parte de teorias metafísicas corretas. D. Lewis não tem intenção de evitar estas consequências e, em vez disso, escreve que:

> «Here there is a rift in our talk of properties, and we simply have two different conceptions. It's not as if we have fixed once and for all, in some perfectly definite and unequivocal way, on the things we call 'the properties', so that now we are ready to enter into debate about such questions as, for instance, whether two of them ever are necessarily coextensive. Rather, we have the word 'property', introduced by way of a varied repertory of ordinary and philosophical uses. The word has thereby become associated with a role in our commonsensical thought and in a variety of philosophical theories. To deserve the name of 'property' is to be suited to play the right theoretical role; or better, to be one of a class of entities which together are suited to play the right role collectively. But it is wrong to speak of *the* role associated with the word 'property', as if it were fully and uncontroversially settled. The conception is in considerable disarray. […]
> There's no point in insisting that this one is the only rightful conception of the properties. Another version of the property role ties the properties more closely to the meanings of their standard names, and to the meanings of the predicates whereby they may be ascribed to things. 'Triangular' means having three angles, 'trilateral' means having three sides. These meanings differ. […] So on this conception of properties, we want to distinguish triangularity from trilaterality, though we never can distinguish their instances. We can put the distinction to use, for instance, in saying that one of the two properties is trivially coextensive with triangularity, whereas the other is non-trivially coextensive triangularity.» (Lewis 1986b: 55-6)

A ideia, então, é que o nosso uso da expressão 'propriedades' – e o mesmo pode ser dito de 'relações' e 'proposições' – não fixa de uma vez para sempre qual é a conceção correta de propriedades que devemos adotar no momento em que as queremos tratar de uma maneira sistemática. Não há maneira de responder, sem controvérsia, à questão de saber se duas propriedades com a mesma extensão em todos os mundos têm de ser

idênticas ou se podem ser diferentes. Numa das conceções disponíveis, aquela que temos vindo a tratar nos parágrafos anteriores, uma propriedade esgota-se no inventário dos seus exemplares atuais e possíveis. Numa outra conceção, no entanto, pretende-se que as propriedades espelhem de uma certa maneira o significado dos predicados usados para as expressar ou dos nomes escolhidos para as denotar. 'Triângulo' e 'trilátero' não significam o mesmo. Logo, nesta conceção, as propriedades *ter três ângulos* e *ter três lados* são distintas. Mais ainda, deixa de haver apenas uma propriedade exemplificada por todos os possíveis e uma propriedade impossível. Na conceção de proposições análoga a esta, pretende-se que duas frases com significados diferentes expressem proposições diferentes, mesmo que recebam o mesmo valor de verdade em todos os mundos. O realismo modal, diz D. Lewis, oferece também os recursos para termos uma ontologia em que se encontrem entidades que desempenham o papel teórico das propriedades, relações e proposições assim concebidas. A ideia é a de construir propriedades (relações e proposições) com uma certa estrutura, através de regras análogas às da sintaxe de uma linguagem, a partir de algumas propriedades mais básicas e sem estrutura, concebidas apenas como conjuntos de objetos possíveis, da mesma maneira que antes. Exemplificando:

> «Let A be the relation of being an angle of; let S be the relation of being a side of. Suppose for simplicity that these can be left as unstructured relations; we could go on to a deeper level of analysis if we like, but that would complicate the construction without showing anything new. Let T be the higher-order unstructured relation which holds between unstructured property F of individuals and an unstructured relation G of individuals iff F is the property of being something which exactly three things bear relation G to. A certain unstructured property is the unique thing which bears T to A, and therefore it is the (unstructured) property of triangularity; it also is the unique thing which bears T to S, and therefore it is the (unstructured) property of trilaterality. Therefore let us take the structured property of triangularity as the pair (T, A), and the structured property of trilaterality as the pair (T, S). Since S and A differ, we have the desired difference between the two pairs that we took to be our two structured properties.» (Lewis 1986b: 56)

Construções deste género podem ser obtidas também para proposições que espelham a estrutura sintática das frases que as expressam. Este é o exemplo de D. Lewis: seja N uma relação que se estabelece apenas entre uma proposição e a sua negação, e seja

P uma proposição qualquer. (A negação de uma proposição concebida como o conjunto de mundos em que é verdadeira tem de ser o conjunto de mundos em que ela é falsa.) A proposição estruturada que corresponde à negação de P pode ser o par ordenado (N, P). (N, (N, P)) é, claro, a proposição estruturada que corresponde à dupla negação de P, que nesta conceção não é mais idêntica à sua equivalente P, como seria se estivéssemos a falar de proposições como conjuntos de mundos. Outras construções permitidas são, por exemplo, os pares ordenados de propriedades e objetos – ou de relações e sequências de objetos. Estes pares ordenados vão corresponder a *proposições singulares*, que são expressas pelas frases predicativas.

Consideramos que permitir uma teoria assim é uma vantagem para o realismo modal apenas se, antes de mais, considerarmos que é útil – ou até indispensável – encontrar um espaço para as propriedades, relações e proposições na nossa ontologia. Em (1983b: 348-51), ainda que a abordar uma questão ligeiramente diferente, D. Lewis aponta algumas aplicações importantes das propriedades e proposições. Uma delas encontra-se no tratamento semântico de algumas frases contendo nomes que, à primeira vista, não podem ser concebidos como denotando objetos particulares, como estas:

(1) O vermelho assemelha-se ao laranja mais do que se assemelha ao azul.

(2) O vermelho é uma cor.

(3) A humildade é uma virtude.

(4) O vermelho é sinal de maturação.

(5) Aquilo que há de comum a todos os que sofrem dor é estarem num ou noutro estado que ocupa o papel causal da dor, presumivelmente não o mesmo em todos os casos.

Ou, ainda, no tratamento de frases que envolvem quantificação de segunda ordem, fazendo com que as propriedades tenham de ser contadas entre os valores de algumas variáveis, por exemplo:

(6) Ele tem as mesmas virtudes que o pai.

(7) Os vestidos eram da mesma cor.

(8) Existem propriedades físicas fundamentais ainda por descobrir.

(9) Características adquiridas nunca são herdadas.

(10) Algumas espécies zoológicas são férteis entre si.

As ideias expressas por algumas destas frases – ou até por todas – podem, como admite D. Lewis, ser também expressas por paráfrases que eliminam o recurso a nomes para propriedades ou à quantificação sobre estas. O problema, no entanto, é que essa tarefa tem de ser feita caso a caso, e por isso ao dispensarmos de uma vez por todas as propriedades, estaríamos a invalidar o esforço de tratar a semântica da nossa linguagem de uma maneira sistemática.

A outra aplicação das propriedades, que inclui também as proposições, é como objetos das atitudes intencionais – crenças, desejos, esperanças, medos, memórias e, possivelmente, episódios de imaginação e de pensamento. Estas atitudes podem ser concebidas, em muitos casos, como relações entre um sujeito e uma proposição. A crença de que a neve é branca, por exemplo, envolve um sujeito e a proposição *que a neve é branca*, identificada na teoria de D. Lewis com o conjunto de mundos em que a neve é branca. Noutros casos, no entanto, as propriedades são mais úteis. Estas questões acerca do conteúdo mental e dos objetos das atitudes serão abordadas mais à frente (*ver* capítulos 3 e 4).

Em (1986b: 174-91), D. Lewis apresenta e discute uma teoria, a que chama ersatzismo mágico, que passa por aceitar as propriedades (e o mesmo vale para as relações e as proposições) como entidades abstratas básicas – irredutíveis a coisas de qualquer outro género – e encontrar entre estas mundos e indivíduos possíveis ersatz. Nesta variedade de ersatzismo, as propriedades não têm uma estrutura interna relevante, e por isso não é em virtude das suas características estruturais que estas conseguem representar cada coisa como sendo desta ou daquela maneira. As propriedades diferem por isso de coisas como frases e imagens. Uma frase representa através da interpretação dos seus constituintes e da maneira como eles são combinados. Uma imagem, por sua vez, pode representar parcialmente por uma interpretação convencional de alguns dos seus elementos, mas também por isomorfismo: pela correspondência entre a estrutura da imagem e a estrutura da situação representada. Nada disso acontece com as propriedades, de acordo com esta variedade de erstazismo.

A teoria pode ser apresentada da seguinte forma. Uma propriedade é *verdadeira de* certas coisas, os seus exemplares, e de mais nenhuma. Em particular, as proposições, que vimos antes poderem ser entendidas como propriedades de mundos, são verdadeiras de certos mundos, e não de outros. De acordo com os proponentes de qualquer variedade de ersatzismo, existe apenas um mundo genuíno, e por isso podemos dizer que uma proposição é verdadeira ou falsa *simpliciter*, consoante o facto de ser ou não verdadeira do único mundo existente.[19] Uma propriedade F representa uma coisa X como sendo de uma certa maneira se e só se, necessariamente, se F é verdadeiro de X, então X é dessa maneira. Em particular, a proposição P representa que tal e tal se e só se, necessariamente, se P é verdadeira, então tal e tal.

Existe uma relação de *implicação* e de *incompatibilidade* entre as propriedades. Uma propriedade F implica uma outra propriedade G se e só se, necessariamente, se F é verdadeira de uma coisa, então G também é verdadeira dessa mesma coisa. Em particular, uma proposição P implica uma proposição Q se só se, necessariamente, se P é verdadeira, então Q também é. Duas propriedades F e G são incompatíveis se e só se não é possível que F e G sejam verdadeiros de uma mesma coisa. Em particular, duas proposições P e Q são incompatíveis se e só se não puderem ser ambas verdadeiras. Com isto, podemos dizer que uma propriedade é *máxima* se e só se é incompatível com qualquer propriedade que não implica. Um indivíduo possível ersatz é uma propriedade máxima, e um mundo ersatz é uma proposição máxima.[20]

A noção de 'verdadeiro de' é aceite como primitiva nesta variedade de ersatzismo, e por isso não se pode exigir ao proponente desta teoria que explique essa noção

---

19 D. Lewis fala em *elementos* que são *selecionados* por certas coisas, e outros que são selecionados apenas pelo único mundo concreto – ou selecionados *simpliciter* –, em vez de falar, como eu fiz aqui, de *propriedades* que são *verdadeiros de* certas coisas, e proposições que são verdadeiras *simpliciter*. O propósito de D. Lewis é usar termos neutros como 'elementos' e 'selecionados', que possam servir para falar de uma variedade de teorias com vocabulário diferente. Em vez de falarem de elementos selecionados por uma coisa, algumas delas falam de maneiras como uma coisa pode ser que são as maneiras como uma coisa é ou de possibilidades que uma coisa realiza (Lewis 1986b: 188). E, ainda, em vez de falarem de elementos selecionados pelo mundo, algumas teorias falam de circunstâncias que obtêm ou de maneiras como as coisas podem ser que são as maneiras como as coisas são (Lewis 1986b: 183).

20 O ersatzismo mágico é uma abordagem próxima àquela que encontramos, por exemplo, em (Plantinga 1974: cap. 4) e (Kripke 1980). Apesar de estes autores não chegarem a dizer exatamente que os mundos e os indivíduos possíveis são entidades possivelmente simples sem estrutura interna relevante, a verdade é que não chegam a explicar exatamente de que modo é que as entidades a que reduzem os possíveis conseguem cumprir o papel que lhes é atribuído.

recorrendo a outras mais básicas. No entanto, D. Lewis considera que se deve exigir que classifique a relação expressa por essa noção como *interna* ou *externa*, e defende que qualquer uma das alternativas é problemática. Uma relação é interna quando depende unicamente do carácter intrínseco dos *relata* – se R é uma relação interna e R(*a*, *b*), quaisquer duplicados possíveis de *a* e *b* também estabelecem a relação R.[21] Vamos assumir que *ser verdadeiro de* é uma relação interna. Isto não pode ser totalmente correto, porque algumas propriedades são extrínsecas – aquelas que um objeto tem por causa da maneira como se relaciona com outras coisas. Por isso, uma propriedade extrínseca não pode ser verdadeira de um objeto em virtude do carácter intrínseco desse objeto. Reformulando, podemos dizer que essa relação é estabelecida unicamente pelas qualidades intrínsecas do mundo inteiro – e já não apenas do objeto relacionado – e da propriedade. Em geral, a relação passa a ser apenas *aproximadamente* interna, mas facilmente se percebe que no caso particular das proposições ela é mesmo interna.

Nós sabemos perfeitamente como é preciso o mundo ser para, por exemplo, ser verdadeira a proposição *que a neve é branca* – mais concretamente, que a neve seja branca. E sabemos como tem de ser o mundo e um objeto para que a propriedade extrínseca *ser a primeira molécula de metano* seja verdadeira dele – mais concretamente, esse objeto tem de ser uma molécula de metano e no mundo não pode haver mais nenhuma molécula de metano a existir em momentos anteriores àquele em que ela se formou, nem nesse mesmo momento. Fica, ainda assim, a questão de saber que características tem de ter uma proposição para que seja verdadeira se e só se a neve é branca, ou uma propriedade para que seja verdadeira apenas da primeira molécula de metano. Não havendo qualquer ilustração do género de características de que estamos à procura, podemos concluir que não é compreensível o que se quer dizer quando se afirma que uma propriedade (ou relação ou proposição) é verdadeira de certos objetos. Esta é uma hipótese de resposta. Tal como não há qualquer problema em dizer que a frase 'A neve é branca' em português tem a característica de representar que a neve é branca, podemos

21 A distinção entre propriedades intrínsecas e extrínsecas será tratada, juntamente com as ideias de propriedades naturais e de duplicação, na secção 3.4. Convém notar que apesar de se poder falar também em relações intrínsecas e extrínsecas, essa classificação não corresponde à classificação das relações em internas e externas.

dizer que as propriedades têm a característica de representar cada coisa como sendo desta ou daquela maneira: nada mais é necessário acrescentar. Para D. Lewis, esta resposta não é adequada porque consiste apenas em dizer que as propriedades ocupam um certo papel teórico em virtude do facto de terem a característica de ocupar esse papel. Falta ainda dizer de que maneira é que as propriedades representam seja o que for. Não há problema nenhum em dizer que a frase 'A neve é branca' representa que a neve é branca, porque se espera que haja uma abordagem que explique como é que isso acontece: ninguém espera que o facto de 'A neve é branca' representar que a neve é branca em vez de, por exemplo, que a relva é verde faça parte do nível mais básico da natureza. Mas suponhamos que podíamos mesmo assumir que há qualidades representacionais fundamentais, instanciadas pelas propriedades. Temos ainda assim o problema de saber como é que é possível nomearmos – ou identificarmos – essas qualidades. As propriedades são entidades abstratas que não estão no espaço-tempo e não se relacionam causalmente nem entre si nem com nada mais. Por isso, não temos qualquer contacto com as qualidades das propriedades, a não ser que essas qualidades sejam exemplificadas por coisas que façam parte da realidade concreta. Tendo em conta que as propriedades, nesta hipótese, não têm uma estrutura interna relevante, as qualidades delas que têm relevância a nível representacional devem poder ser exemplificadas por átomos mereológicos (coisas simples, sem partes), e os exemplos disponíveis de propriedades instanciadas por coisas simples no mundo concreto – coisas como massa, carga, a cor e o sabor dos quarks, entre outras – não são tão vastas quanto as entidades representacionais necessárias.

Agora vamos assumir que a relação *ser verdadeiro de* é uma relação externa. Uma relação externa, como a distância entre dois pontos, não depende do carácter intrínseco dos *relata*, mas apenas do carácter intrínseco do agregado deles. Nesta hipótese, são completamente irrelevantes as qualidades intrínsecas das propriedades: importa apenas a disposição de cada uma numa teia de relações com o mundo e com as suas partes. Existem, ainda assim, dois problemas com esta hipótese. O primeiro é idêntico ao problema que antes se considerou acerca da identificação das qualidades representacionais, mas aplicada agora diretamente à relação de ser verdadeiro de uma coisa: se esta é instanciada por pares de objetos em que um deles está fora dos limites do

espaço e do tempo e não exerce qualquer influência causal sobre aquilo que ocorre dentro desses limites, como é que alguma vez podemos estabelecer contacto com essa relação de modo a pensá-la ou nomeá-la? O outro problema é específico a esta segunda alternativa. A relação de ser verdadeiro de uma coisa é uma relação *modal*. Uma propriedade que representa uma coisa como sendo branca necessariamente é verdadeiro de uma coisa branca: esta relação não é acidental. Por rejeitar conexões necessárias entre entidades distintas, D. Lewis considera que esta ideia é inaceitável: «It seems to be one fact that somewhere within the concrete world, a donkey talks; and an entirely independent fact that the concrete world enters into a certain external relation with this element and not with that. What stops it from going the other way? Why can´t anything coexist with anything here: any pattern of goings-on within the concrete world, and any pattern of external relations of the concrete world to the abstract simples?» (Lewis 1986b: 180)[22]

1.6 – Análise redutiva da modalidade

Vimos antes que D. Lewis concebe um mundo como uma soma de indivíduos conectados espáciotemporalmente entre si e com nada mais, e que trata a modalidade *de dicto* como quantificação sobre mundos. Assim, com esta interpretação, quando se faz a afirmação *de dicto* de que é possível haver plantas falantes, o que se está a dizer é que

[22] Em (1986: 202-10), Peter van Inwagen defende que tem de haver alguma coisa de errado com este argumento contra as propriedades, relações e proposições, assim como são concebidas pelo ersatzismo mágico, porque, afirma, podemos construir um argumento parecido para provar que a relação de ser membro de um conjunto também não é inteligível. Esta é uma ideia inaceitável, pela consequência de que não entendemos o resto da matemática que se fundamenta na teoria dos conjuntos, e é especialmente inaceitável para D. Lewis, que baseia toda a sua ontologia em objetos concretos possíveis e contruções conjuntistas a partir destes, como testemunha aquilo que tenho vindo a dizer neste capítulo. Mais tarde, em "Mathematics is Megethology" (1993), D. Lewis vai reconhecer este problema. Mas, ao contrário de Peter van Inwagen, que adota a conceção do ersatzismo mágico, vai manter a sua crítica a esta teoria e formula, em vez disso, uma ontologia estruturalista da teoria dos conjuntos. Não vou poder abordar esta proposta aqui, mas a ideia geral é a de que não existe uma única coisa que é o conjunto dos indivíduos X e Y, mas uma coisa que está associada a X e Y através de uma certa função, e de acordo com essa função pode ser chamada o conjunto de X e Y. Noutra função essa mesma coisa pode ser o conjunto de Y e Z, por exemplo. Os teoremas da teoria dos conjuntos, e o nosso discurso comum sobre conjuntos, têm então de ser entendidos como falando universalmente de todas essas funções. Gideon Rosen (2015: 392-95) nota que a aplicação de uma estratégia estruturalista às propriedades, relações e proposições pode também fazer com que o ersatzismo mágico escape aos problemas que D. Lewis apontou.

algumas dessas somas chamadas 'mundos possíveis' têm plantas falantes como parte. Desapareceram nesta paráfrase os conceitos modais: não foi preciso falar de possibilidade e necessidade para dizer o que é um mundo nem para dizer o que é verdade em cada um deles. O mesmo acontece no caso da modalidade *de re*. Dizer que Bertrand Russell é essencialmente um humano é o mesmo que dizer que em qualquer uma dessas somas chamadas 'mundos possíveis', se nelas há uma ou mais do que uma contraparte de Russell, então essas contrapartes também são humanas. A relação de contraparte, por sua vez, é uma relação de semelhança entre coisas atuais e meramente possíveis. Em todo este processo podemos falar daquilo que é essencial e acidental a uma coisa sem recorrer a conceitos modais. Chega-se assim a uma *análise redutiva* da modalidade – a uma explicação completa dos conceitos modais de necessidade e possibilidade *de dicto* e *de re* que não utiliza conceitos deste género. Deste modo, a modalidade deixa de fazer parte do inventário das noções primitivas que entram na nossa teoria total do mundo, que pode agora ser expressa numa linguagem puramente extensional, utilizando apenas os recursos formais da teoria da quantificação com identidade, da mereologia e da teoria dos conjuntos, além de uma série de conceitos extensionais como os de semelhança, de relação espáciotemporal, entre outros. A modalidade não é abandonada, mas apenas analisada. A teoria total do mundo é de alguma maneira uma metalinguagem que atribui condições de verdade às frases que usamos, abertamente, acerca do tema da modalidade. Apesar de as condições de verdade estarem especificadas sem apelar a conceitos modais, elas estabelecem quando é que uma afirmação abertamente modal é verdadeira e falsa, e por isso fazem deste assunto algo real e genuíno.

Esta análise da modalidade em termos de conceitos puramente extensionais é, como refere Scott Soames, mais um dos aspetos em que D. Lewis recebe a influência de W. V. Quine:

> «Here we see what appears to have been the enduring influence of David's teacher, Quine, the great champion of naturalism and extensionalism, and the uncompromising scourge of the modalities. The underlying philosophical purpose of modal realism and counterpart theory was to reduce an intensional object-language to a purely extensional semantic metalanguage, in the service of an antecedently desired conception of reality. Whereas Quine taught that vindicating naturalism and extensionalism required eliminating intensional facts and rejecting

> intensional constructions, his student, David Lewis, tried to show that intensional facts are just a species of extensional facts, and that intensional constructions in language are no threat to the integrity of an austere, naturalistic vision of reality.» (Soames 2015: 83)

Normalmente fala-se de um contexto extensional numa frase como um contexto em que termos com a mesma *extensão* podem ser substituídos mantendo estável o valor de verdade da frase inteira, e fala-se de uma linguagem extensional como uma linguagem em que todos os contextos são extensionais. Aquilo que conta como uma extensão adequada não é o mesmo em cada categoria de expressões. A extensão de um predicado é o conjunto de coisas das quais o predicado é verdadeiro (podem ser sequências de coisas se o predicado é poliádico), a extensão de um termo singular (como um nome ou uma descrição definida) é o seu referente e, por fim, a extensão de uma frase é o valor de verdade que possui. As atribuições de extensão a um termo devem ser relativas a um contexto – e, em particular, são afetadas por questões de facto contingentes. As extensões fornecem relativamente pouca informação acerca das propriedades semânticas de um termo, a não ser que se fale também da maneira como elas são variáveis entre contextos atuais e possíveis. As *intensões* são coisas que capturam essa informação: são funções que associam a cada contexto uma extensão. Esquecendo qualquer outra variação contextual a não ser a localização no espaço lógico, é conveniente identificar as intensões na teoria de D. Lewis com funções de mundos para extensões. Mas desta maneira perde-se a possibilidade de dizer que a teoria total do mundo proposta por D. Lewis é uma teoria extensional. Da mesma maneira que os operadores modais aplicados criam um contexto intensional na frase a que são aplicados, o mesmo vai acontecer com a quantificação sobre mundos que analisa esse idioma: 'criatura com rins' e 'criatura com coração' têm a mesma extensão no mundo atual e uma diferente intensão, mas mesmo assim não são substituíveis nem numa frase como 'É possível que as criaturas com rins não tenham coração' nem em 'Em alguns mundos, as criaturas com rins não têm coração' (que é, esta última, a frase que de acordo com a teoria de D. Lewis expressa a ideia a um nível mais básico).

Este problema parece-me ser meramente uma questão de terminologia. É conveniente, para analisar semanticamente a nossa linguagem que fala acerca do nosso

mundo como o atual e dos restantes como meramente possíveis, dizer que a extensão de um termo é aquela que é atribuída no mundo atual. Mas a teoria total do mundo proposta por D. Lewis é completamente neutra relativamente à nossa localização no espaço lógico – é completamente indiferente para as verdades que essa teoria exprime que estejamos neste mundo ou noutro qualquer: ela permite-nos de facto uma visão sobre tudo aquilo que nós descreveríamos como a atualidade e, ainda, sobre toda a mera possibilidade. Deste modo, pode ser conveniente para analisar semanticamente esta teoria que a extensão dos termos que nela ocorrem seja atribuída tendo em conta tudo o que existe no espaço lógico e não apenas no mundo atual. Nesse caso, 'criatura com rins' e 'criatura com coração' não são mais termos extensionalmente equivalentes, porque muitas criaturas com rins espalhadas pelos mundos não são criaturas com coração. Ainda assim, se quisermos continuar a reservar o termo 'extensão' para falar do aspeto semântico de um termo que é relativo a cada mundo, continuamos a poder afirmar que, a um nível fundamental, a lógica adequada para a teoria total do mundo que D. Lewis adota é a mesma que aquela que é adequada para lidar com as linguagens que não envolvem contextos intensionais. Esta ideia é suficiente para dizer que basicamente a realidade não está partida em factos que podem ser expressos através dos recursos de uma lógica extensional e aqueles que exigem a intervenção de uma lógica intensional.

Em (1986b: cap. 3), D. Lewis apresenta e discute três variedades de ersatzismo (mágico, pictórico e linguístico), e encontra nas três a desvantagem, relativamente ao realismo modal genuíno, de terem de assumir a modalidade como primitiva, de modo (1) a caracterizarem em que consiste um mundo e indivíduo possível ersatz ou (2) a explicarem como é que cada mundo e indivíduo ersatz representa. Comecemos pelo ersatzismo mágico. Como vimos antes (secção 1.5.5), nesta teoria os mundos e os indivíduos possíveis ersatz eram tratados como elementos abstratos simples de um certo tipo – mais concretamente, um indivíduo era um elemento máximo, e os mundos eram os casos particulares de indivíduos que só podiam ser verdadeiros do mundo inteiro. A modalidade tem de estar envolvida na definição de 'máximo'. Um elemento é máximo se e só se for incompatível com todos os elementos que não implica. E, por fim, E implica F se e só se, *necessariamente*, se E é verdadeiro de uma coisa, então F também é

verdadeiro dessa coisa. E é incompatível com F se e só se não é *possível* que E e F sejam ambos verdadeiros de uma mesma coisa.

No erstazismo pictórico, por sua vez, trata-se um mundo ersatz como uma imagem, ainda que abstrata, e um indivíduo possível ersatz como uma parte qualquer de uma dessas imagens. Um desses mundos ersatz representa corretamente o mundo atual por lhe ser *integralmente* isomórfico[23] – i. e., a cada parte desse mundo ersatz corresponde uma parte do mundo atual que exemplifica exatamente as mesmas propriedades e que estabelece as mesmas relações com as restantes partes do mundo atual que a parte correspondente do mundo ersatz estabelece com as restantes partes desse mundo ersatz. Dito de um outro modo, esse mundo ersatz e o mundo atual apresentam o mesmo padrão de instanciação de propriedades: são estruturalmente idênticos e cada parte das suas estruturas é intrinsecamente semelhante. Os restantes mundos ersatz também representam o mundo atual como sendo de uma certa maneira através da estrutura que apresentam, mas representam-no falsamente: as relações e as propriedades instanciadas pelas suas partes não são as mesmas que as relações e propriedades instanciadas pelas partes do mundo atual. Devemos ainda acrescentar que estes mundos ersatz representam o mundo atual em completo detalhe (Lewis 1986b: 165-67). Queremos dizer que certas coisas são

23 Um mundo ersatz é uma imagem diferente das imagens comuns que costumamos fazer e usar. Como D. Lewis nota (1986b: 166), estas últimas representam apenas *parcialmente* – não integralmente – por isomorfismo. Imaginemos um desenho que representa um gato num tapete – um gato em concreto, aquele que a pessoa que o desenhou pode dizer, apontando para ele, «É *este* gato que representei no meu desenho». (O exemplo é do próprio D. Lewis.) Nesse desenho, aparece uma linha, que representa um tapete, acima da qual aparece um esboço de contornos característicos de um mamífero, que representa, claro, o gato. Num certo sentido, é correto dizer (1) que o desenho representa *aquele* gato como estando em cima de um tapete e (2) fá-lo por isomorfismo, ou seja, apresenta certos elementos – a linha e o esboço em forma de mamífero – como estando numa relação semelhante àquela que é estabelecida entre o gato e um tapete (ou àquela que se estabeleceria caso o gato estivesse em cima de um tapete, se a situação for contrafactual). Mas esse isomorfismo é, mesmo na melhor das hipóteses, limitado. Por exemplo, as partes do gato e as partes do tapete estão muito mais próximas no desenho no que na realidade, apesar da imagem não as representar como estando assim tão próximas; também só aparecem no desenho as partes do gato que são vistas desde uma certa perspetiva, sem que com isso se diga que o desenhe o gato como incompleto, entre outras coisas. A lição que daqui se deve retirar é esta: se uma imagem tivesse de ser integralmente isomórfica àquilo que representa, então o desenho *daquele* gato no tapete não seria uma imagem *daquele* – nem de nenhum outro! – gato. Por isso, parte dos factos que fazem com que a imagem seja acerca daquele gato têm de ser convencionais. Não queremos que isto aconteça com os mundos ersatz, segundo D. Lewis, porque nesse caso estaríamos ainda numa posição muito próxima à do ersatzismo linguístico. (As expressões linguísticas adquirem significado por convenção.) O afastamento entre estes dois tipos de ersatzismo é o que faz com que as duas posições tenham vantagens e desvantagens distintas, as quais D. Lewis pretende tratar separadamente.

verdade relativamente a alguns destes mundos ersatz e falsas relativamente a outros. É verdade relativamente a um destes mundos ersatz aquilo que ele representa como sendo verdade. Acabámos de ver que uma imagem representa por isomorfismo e, por isso, concebendo os mundos ersatz como imagens, vamos dizer que um deles representa, por exemplo, que existe um gato por ter uma parte que é isomórfica a um gato. Até agora não foi preciso apelar a conceitos modais para explicar o que é um mundo ersatz na variedade pictórica, nem como eles conseguem representar a realidade como sendo de uma certa maneira. Mas só conseguimos fazer isso porque escolhemos um exemplo certo. Tentemos dizer como tem de ser um mundo ersatz para representar que existe um unicórnio. A resposta que demos no caso da existência de um gato já não funciona, tendo em conta que não há qualquer unicórnio ao qual uma parte desse mundo ersatz seja isomórfica. A alternativa é dizer que o mundo ersatz representa que existe um unicórnio porque contém uma parte que (1) *poderia* ter sido isomórfica a um unicórnio, (2) *seria* isomórfica a um deles se o mundo fosse diferente e contivesse um unicórnio com partes que instanciam certas qualidades intrínsecas e estão estruturalmente arranjadas de uma determinada maneira, como as partes da parte do mundo que ersatz que pretendemos que represente um unicórnio, e (3) não *poderia* ter sido isomórfica a nenhuma coisa exceto um unicórnio. Acabamos de utilizar os conceitos modais, sem qualquer hipótese de os analisar redutivamente. Se pretendermos analisar a necessidade e a possibilidade em termos de quantificação sobre mundos ersatz, vamos ter de dizer que possivelmente existem unicórnios se e só se em algum dos mundos ersatz é verdade que existem unicórnios. Acabamos de ver que para dizer o que é verdade num mundo ersatz temos de falar daquilo que ele representa como sendo verdade, e que para dizer como é que um mundo ersatz representa que existe um unicórnio precisamos de falar naquilo que é ou não possível. A análise pretendida revela-se circular e, por isso mesmo, inadequada (Lewis 1986b: 168).

Finalmente, no ersatzismo linguístico os mundos ersatz são coisas como contos ou teorias que descrevem completamente o mundo concreto. Assumindo que os contos e as teorias podem ser identificados com conjuntos de frases (interpretadas), a melhor maneira de conceber um mundo ersatz é como um conjunto de frases ao mesmo tempo máximo e consistente (Lewis 1986b: 142). Tal como acontecia nas outras variedades de

ersatzismo, é verdade em cada mundo ersatz aquilo que cada um representa como sendo verdade. Desta vez, a representação acontece através das frases que pertencem aos mundos ersatz. Um mundo ersatz representa que existem unicórnios se e só se (1) esse mundo tem como elemento a frase (em português) 'Existem unicórnios' ou outra com as mesmas condições de verdade, ou (2) um subconjunto desse mundo ersatz implica que existem unicórnios (por exemplo, esse mundo contém frases que descrevem uma distribuição de propriedades químicas que envolve a presença de um unicórnio). O modo de representação (1) é explícito, enquanto que (2) é implícito.[24] Tudo o que foi dito até agora para os mundos ersatz é aplicável aos indivíduos possíveis ersatz ainda que com uma leve modificação: as frases que pertencem a cada um destes indivíduos devem ser frases abertas com uma única variável que ocorre livre. Assim, em vez de serem descrições verdadeiras ou falsas (do mundo inteiro), são descrições verdadeiras ou falsas *de* certas coisas (Lewis 1986b: 149).

Na apresentação desta teoria, utilizamos implicitamente conceitos modais em duas ocasiões. Em primeiro lugar, vimos que um mundo ersatz tem de ser um conjunto *consistente* e *máximo* de frases. É máximo um conjunto consistente de frases que deixe de ser consistente quando lhe acrescentamos uma outra frase que não é por ele implicada. Ficamos com as noções de implicação e consistência ainda por analisar – ambas, à primeira vista, noções modais. Comecemos por esta última: «[...] in order to say which things of the right nature – which sets of sentences of the worldmaking language – are the ersatz worlds, we need to distinguish the consistent ones. That is *prima facie* a modal distinction: a set of sentences is consistent iff those sentences, as interpreted, *could* all be true together.» (Lewis 1986b: 151) Em segundo lugar, vimos que um mundo ersatz representa implicitamente que tal e tal se e só se um subconjunto desse mundo *implica* que tal e tal. A noção de implicação aparece aqui outra vez. «[T]his implication is *prima facie* modal: a set of sentences imply that so-and-so iff those sentences, as interpreted, *could* not be true together unless it were also true that so-and-so; in other words, if it is

[24] Claramente, o primeiro modo de representação pode ser subsumido no último. Quando uma frase afirma explicitamente que tal e tal, a frase implica que tal e tal. Por isso, podíamos antes dizer apenas isto: um mundo ersatz representa que tal e tal se e só se nele estiver contido um conjunto de frases que implique que tal e tal.

necessary that if those sentences are all true together, then so-and-so.» (Lewis 1986b: 151)

Em vez de utilizarmos conceitos modais, por que razão não analisamos a consistência e a implicação em termos semânticos? Um conjunto de frases seria assim consistente se e só se há uma reinterpretação do vocabulário não-lógico que ocorre nessas frases que torna todas elas conjuntamente verdadeiras. E, de uma maneira semelhante, um conjunto de frases K implica uma outra frase ϕ se e só se não existe qualquer reinterpretação do vocabulário não-lógico que ocorre nas frases de K e em ϕ que torne verdadeiras todas as frases em K e ao mesmo tempo falsa ϕ. Esta análise captura corretamente as noções de consistência e implicação estritamente lógicas, mas não as noções mais gerais que precisamos para identificar os mundos e indivíduos possíveis ersatz. Reparemos apenas que, se estas fossem adequadas, haveria um mundo ersatz de acordo com o qual uma pessoa é ao mesmo tempo solteira e casada, e outro de acordo com o qual existem moléculas de metano sem haver átomos de carbono e de hidrogénio, entre muitos outros casos assim. É impossível que o mundo concreto seja da maneira como estes supostos mundos possíveis ersatz o descrevem, e por isso eles devem ser categorizados, no máximo, como mundos *impossíveis* ersatz. Continuamos a precisar de falar em possibilidade e necessidade para distinguirmos de entre todos os mundos ersatz aqueles que são os possíveis.

Há efetivamente uma estratégia para evitar a modalidade primitiva neste contexto, ainda que, muito provavelmente, seja impossível de concretizar na prática, e pareça ainda um projeto de alguma maneira inadequado. A ideia é, em parte, tentar analisar a consistência e a implicação em termos sintáticos, em vez de modais, utilizando um sistema de regras formais de dedução (talvez com alguns axiomas lógicos). Mas esta é apenas uma parte do projeto, tendo em conta que através apenas destas regras (e dos axiomas lógicos) não vamos chegar mais longe do que a uma análise da consistência e implicação estritamente lógicas, que acabamos de ver serem inadequadas para o propósito de teorizar acerca dos mundos possíveis ersatz. A outra parte passa por especificar certas frases de cada linguagem com que se pretende construir mundos ersatz que não sejam logicamente verdadeiras, e categoriza-las depois como *axiomas*. A inconsistência mais

lata pode acontecer entre duas frases logicamente consistentes porque uma delas implica uma terceira frase que é logicamente inconsistente com a outra, ou porque ambas são formadas por predicados que de alguma maneira têm uma qualquer conexão modal (por exemplo, de implicação, incompatibilidade ou equivalência) que não é detetável por meios lógicos. Essa conexão, na maior parte das vezes, acontece por causa do significado dos predicados, como é o caso com 'casado' e 'solteiro', mas não devemos por de lado imediatamente a ideia de outros fatores estarem presentes. Imaginemos que não é possível uma partícula ter ao mesmo tempo carga positiva e carga negativa. Este seria o exemplo de um facto modal que não decorre do significado dos termos usados e que levaria a uma conexão de incompatibilidade entre dois predicados. (Se as leis da natureza forem necessárias, temos aí um outro exemplo.) Tendo isto em conta, o papel dos axiomas neste projeto é (1) dizerem que não é conjuntamente verdade de nenhuma coisa aquilo que é expresso por certos predicados que estabelecem uma relação de incompatibilidade que não é logicamente detetável,[25] e (2) dizerem que *se* for verdade conjuntamente aquilo que é expresso por certas frases num conjunto K que implica uma frase ϕ, ainda que não logicamente, *então* também é verdade aquilo que é expresso pela frase ϕ. É importante acrescentar ainda que os axiomas, para cumprirem a tarefa que lhes é atribuída, devem falar das ideias expressas pelos predicados e pelas frases de que tratam utilizando precisamente esses predicados e essas frases.[26] Ao cumprirem este papel, os axiomas

25 Os casos de implicação e equivalência são explicáveis a partir da incompatibilidade. Um predicado F implica um outro G se e só se F é incompatível com não-G. E, ainda, F é necessariamente equivalente a G se e só se F implica G e G implica F.

26 Não seria adequado que os axiomas expressassem aquilo que é pretendido acerca de certos predicados e frases utilizando predicados e frases equivalentes, mesmo que pertencentes à mesma linguagem, nem sequer usando nomes que denotem esses predicados e frases e falando das relações entre a verdade de cada um deles de uma maneira metalinguística, por ascensão sintática. Utilizar frases equivalentes iria impedir que os axiomas se relacionassem adequadamente – i. e., sintaticamente – com as frases acerca das quais se pretende que digam algo, a não ser que existissem outros axiomas que fizessem essa conexão, o que tornaria as coisas muito mais complicadas. As coisas também seriam muito mais complicadas se os axiomas fossem formulados numa metalinguagem que trata das propriedades semânticas dos predicados e das frases da linguagem com que se constroem os mundos ersatz, exigindo ainda axiomas que relacionassem a metalinguagem com a linguagem. Como não são formulados numa metalinguagem, estritamente falando não são acerca de predicados e frases da linguagem dos mundos ersatz, mas obviamente é pensando metalinguisticamente que eles são especificados e, em todo o caso, é útil descrever a sua tarefa como de algum modo metalinguística. Exemplificando, não devemos ter como axioma a frase "Não há nenhuma coisa da qual sejam verdadeiros tanto o predicado 'casado' como o predicado 'solteiro'", mas simplesmente 'Nenhum solteiro é casado'.

carregam a informação que faltava nas regras de dedução para que consigamos chegar até às noções latas de consistência e implicação a partir das estritamente lógicas. (Com esta especificação de frases axiomáticas, também podemos recuperar o método semântico que foi rejeitado no parágrafo anterior. Às definições de consistência e de implicação aí oferecidas acrescenta-se apenas que as reinterpretações relevantes são apenas aquelas que tornam verdadeiros todos os axiomas especificados.) Um dos problemas que D. Lewis encontra nesta estratégia é que, mesmo existindo os axiomas pretendidos, não é de modo nenhum fácil especifica-los a não ser que se utilizem conceitos modais. E, nesse caso, voltamos à modalidade primitiva. É praticamente impossível, por exemplo, arranjar uma maneira de especificar todos os axiomas que digam o que é verdade caso se verifique a presença de partículas fundamentais com certas qualidades intrínsecas e arranjadas espáciotemporalmente de uma determinada maneira:

> «We need *connecting* axioms: conditionals to the effect that *if* – here follows a very long, perhaps infinitary, description of the arrangement and properties of the point particles – *then* there is a talking donkey. If we are not to get the facts of modality wrong, such axioms had better be written into the definitions of consistency and implication.
> That would be impossible in practice, of course. Maybe the axioms would be infinite both in length and in number; I would suppose not, but at any rate it is safe to say that nobody could produce them. So no linguistic ersatzer can complete his theory. Unless, of course, he does the obvious thing: declares wholesale that among all conditionals with local antecedents and global consequents, exactly those shall be axioms that are necessarily true.» (Lewis 1986b: 156-7)

E, D. Lewis ainda acrescenta, mesmo que contra todas as expetativas um teórico do realismo modal ersatz conseguisse escrever todos os axiomas requeridos para que a teoria explique adequadamente a modalidade, ou consiga pelo menos especificá-los recursivamente, ainda assim há alguma coisa de errado na ideia de que temos de fornecer uma análise de factos globais do mundo em termos de factos locais acerca da distribuição de propriedades e relações entre as partículas fundamentais para que tenhamos uma análise adequada da modalidade (Lewis 1986b: 157). Teríamos de dizer como é que o mundo tem de ser a um nível básico para haver, por exemplo, moléculas, células, gafanhotos, árvores, estrelas e buracos negros, de modo a pudermos dizer o que é

necessário e possível. Esta ideia parece efetivamente absurda – e parece inverter a ordem natural de como a teorização acerca destas matérias deve decorrer.

Há, ainda assim, uma característica importante do ersatzismo linguístico. De entre as variedades de realismo modal ersatz que considerámos, esta é a única que é compatível qualitativamente com a ontologia de D. Lewis: aceitar esta teoria não nos compromete com mais do que particulares concretos e conjuntos. Mundos e indivíduos ersatz são conjuntos de frases – fechadas ou abertas, respetivamente. Uma frase pode ser identificada com a sequência de palavras simples a partir das quais é construída – ou, de uma maneira mais estruturada, com a sequência das expressões que são os seus constituintes imediatos, que por sua vez são identificados também com a sequência dos seus constituintes imediatos, até chegarmos às palavras simples, que são apenas conjuntos das suas inscrições concretas (Lewis 1986b: 142-3). Ainda que qualitativamente semelhante à de D. Lewis, a ontologia do ersatzismo linguístico é ainda mais parcimoniosa a nível quantitativo, porque dispensa todos os objetos concretos meramente possíveis. Temos assim uma teoria que, tendo de assumir a modalidade como primitiva, tem uma ideologia mais complicada que a do realismo modal genuíno, mas uma ontologia com o mesmo número reduzido de categorias básicas de entidades. Ideologicamente (i. e., ao nível dos conceitos que utiliza), no entanto, esta teoria tem de entender a modalidade como primitiva, ao contrário do que acontece com o realismo modal genuíno.

1.7 – Princípio de plenitude e recombinação

Para que seja correta a paráfrase do nosso discurso modal em termos de quantificação sobre mundos possíveis assim como D. Lewis os entende, como universos isolados, tem de ser verdade que absolutamente cada maneira como um mundo pode ser é a maneira como um dos mundos efetivamente é. Esta ideia, o *princípio de plenitude*, deve por isso ser incorporada na teoria de D. Lewis, de modo a garantir que existe a completude e abundância do espaço lógico necessárias a que os mundos desempenhem adequadamente o papel teórico que lhes está associado. No entanto, como vimos, D. Lewis pretende que falar de mundos forneça uma análise dos conceitos de possibilidade

e necessidade, e não apenas uma maneira equivalente de falar das mesmas coisas, e sendo assim não há realmente qualquer conteúdo em dizer que cada maneira como o mundo pode ser é a maneira como pelo menos um dos mundos efetivamente é. Dizer isso quer apenas dizer que a maneira como um dos mundos é identifica-se com a maneira como pelo menos um dos mundos é – uma verdade lógica trivial. A 'maneira como uma coisa é' ou 'a maneira como uma coisa pode ser' parece ser uma propriedade maximal que uma coisa tem ou pode ter. No enquadramento teórico do realismo modal, uma 'maneira como um mundo pode ser' deve identificar-se com um conjunto que tenha como membro um único mundo (ou talvez vários mundos indiscerníveis). A outra alternativa é dizer que uma 'maneira como um mundo pode ser' é basicamente um mundo concreto. (D. Lewis não opta por nenhuma dessas alternativas, e considera até essa questão irrelevante (Lewis 1986b: 86-7).) Em qualquer um dos casos, o princípio de plenitude é verdadeiro mesmo que exista apenas um mundo, dezassete mundos, ou nenhum.

O princípio de plenitude tem por isso de ser reformulado, ou então a modalidade deve ser tomada como primitiva, e deixa de existir a vantagem decisiva do realismo modal genuíno relativamente ao realismo modal ersatz. Uma hipótese que D. Lewis considera é que cada maneira que nós *pensamos* que um mundo pode ser é a maneira como um dos mundos efetivamente é. Apesar de este ser um princípio de plenitude genuíno, é inaceitável, de acordo com D. Lewis, tendo em conta que aquilo que se pretende é uma teoria que permita uma análise redutiva da modalidade, mas não uma teoria acerca daquilo que é ou não possível (Lewis 1986b: 87).

A sua proposta é a de incorporar no realismo modal um *princípio de recombinação*, que afirma que colocar lado a lado partes de vários mundos produz um outro mundo. Se há, por exemplo, um dragão num mundo e um unicórnio noutro, então tem de haver um mundo em que há um dragão e um unicórnio contíguos no espaço-tempo. A ideia é que qualquer coisa pode existir com qualquer outra coisa, não havendo conexões necessárias entre coisas distintas. Isso quer dizer, portanto, que existe um mundo que junta os habitantes de dois outros mundos. É preciso ter em atenção, ainda assim, que D. Lewis não pretende analisar esta ideia através dos recursos da teoria das contrapartes. Quando diz que um dragão e um unicórnio que habitam dois mundos diferentes podem

existir lado a lado não pretende dizer, neste contexto, que uma contraparte desse dragão e uma contraparte desse unicórnio existem lado a lado num mundo possível. E não pretende dizê-lo porque a relação de contraparte envolve também uma semelhança a nível de propriedades extrínsecas, como a origem de cada coisa. Se, por exemplo, todas as contrapartes de um unicórnio tiverem de habitar em mundos que se pareçam com o mundo do unicórnio, pelo menos no que diz respeito à história que desencadeia o aparecimento dessa criatura, então não existe um mundo em que o unicórnio e o dragão existam em conjunto e nada mais exista além deles, nem existe um mundo em que o dragão e o unicórnio sejam as primeiras coisas a aparecerem, como é pretendido pelo princípio de recombinação. A lição que se retira daqui é a que os aspetos extrínsecos devem ser irrelevantes neste contexto, e apenas deve ser tido em conta o carácter intrínseco das coisas que se diz que podem coexistir. Assim, em vez de se falar em contrapartes deve falar-se em *duplicados*: o princípio de recombinação afirma que um duplicado de qualquer coisa pode coexistir com um duplicado de qualquer outra coisa (Lewis 1986b: 88-9).

Alguém poderia pensar que o princípio de recombinação permite gerar todo o espaço lógico a partir daquilo que existe no mundo atual, mas isto não é verdade. Entre os indivíduos possíveis existem alguns que, mesmo não sendo duplicados de nenhum indivíduo atual, são divisíveis em partes que são duplicados de partes do mundo atual. Outros, no entanto, chamados indivíduos *estranhos*, não têm nenhuma parte com o mesmo carácter intrínseco que um indivíduo atual. Também existem propriedades estranhas, que não são instanciadas por nenhuma parte do mundo atual, nem podem ser construídas como propriedades estruturais ou de uma maneira verofuncional a partir das propriedades de coisas atuais. Um indivíduo que instancia propriedades estranhas é, também ele, estranho – os mundos que têm partes que instanciam propriedades estranhas são também estranhos. Mas nem todos os indivíduos estranhos instanciam propriedades estranhas. Suponhamos que noutros mundos existem partículas com, ao mesmo tempo, carga positiva e carga negativa. Esses indivíduos não são duplicados de nenhum indivíduo atual, mas nenhuma das suas propriedades é estranha. Estamos agora em condições de compreender porque é errado dizer que todo o espaço lógico pode ser gerado a partir do

mundo atual: ao recombinarmos coisas atuais não conseguimos gerar indivíduos estranhos. Se pensarmos que a cada propriedade corresponde um universal ou um tropo, talvez pudéssemos gerar os indivíduos estranhos que não têm propriedades estranhas, mas não todos os restantes.

O propósito do princípio de recombinação não é, no entanto, o de gerar os restantes mundos possíveis, mas garantir apenas que não existem falhas no espaço lógico que impeçam os mundos de cumprir a sua tarefa na análise redutiva da modalidade. Se existirem mundos estranhos, garante-se que as coisas que neles existem estão sujeitas ao princípio de recombinação; se não existirem, então não é possível haver coisas estranhas. Qualquer uma destas alternativas é aceitável. Por isso não é um problema para o realismo modal genuíno que a recombinação das coisas atuais permita obter apenas uma parte de todos os mundos. Esta é, ainda assim, uma consequência que parece inaceitável para o ersatzismo linguístico. De acordo com esta teoria, os mundos representam linguisticamente, e para que alguns deles representem a existência de indivíduos com propriedades estranhas é exigido que existam essas propriedades estranhas no mundo atual, para que certas palavras as possam referir. Mas, por hipótese, não existe mais nada para além daquilo que é atual – e no mundo atual não existem obviamente propriedades estranhas a esse mundo.

É inaceitável responder a esta objeção dizendo que não é possível haver propriedades estranhas ao mundo atual. Não é razoável negar que existam mundos mais ricos qualitativamente que o nosso, mesmo que isso, apesar de implausível, até seja correto. Para se perceber porquê pense-se que, apesar de por hipótese o nosso mundo ter um exemplar de cada propriedade, ainda assim existem outros mundos que contêm menos propriedades que o nosso. Suponhamos que um desses mundos mais empobrecidos era o atual, e imaginemos que nós existíamos nesse mundo e estávamos a discutir metafísica da modalidade. Para defender uma posição como o ersatzismo linguístico, um de nós afirmava nesse mundo que talvez não existisse qualquer mundo possível em que as propriedades instanciadas fossem diferentes das desse mundo. Mas atualmente nós sabemos que isso está errado, e não temos qualquer garantia de que não estamos numa situação semelhante à dessa pessoa imaginária. Repare-se ainda que se o ersatzismo

linguístico fosse a forma correta de pensar sobre a modalidade, então nesse mundo empobrecido aquele que é o nosso mundo nem sequer seria um mundo possível.

Uma outra forma de resolver este problema passa por lembrar que nomear um objeto ou uma propriedade não é a única maneira de se falar acerca deles: é possível fazê-lo também por descrição. Desta forma, talvez não precisemos de ter expressões que correspondam às propriedades estranhas, pelo que os predicados que expressam propriedades atuais podem ser suficientes para formarmos descrições que correspondem a essas propriedades e que, deste modo, nos permitem construir todos os mundos ersatz em falta. As frases de Ramsey podem ser uma ferramenta útil para a construção dos mundos. Imaginemos que a frase

$$\mathrm{A}(a_1, \ldots, a_n).$$

É a conjunção de todas as frases que pertencem ao único mundo ersatz atual. Assim sendo, a frase descreve completamente o mundo concreto e, consequentemente, descreve integralmente o papel que as propriedades e relações exemplificadas nesse mundo – nomeadas por 'a_1', …, 'a_n' – aí ocupam. A partir desta frase, construímos a correspondente frase de Ramsey substituindo os nomes para as propriedades e relações por variáveis livres, acrescentando depois vários quantificadores existenciais no início da frase de modo a ligar todas as variáveis. Depois desse processo, temos então esta frase de Ramsey:

$$\exists x_1, \ldots, x_n \, \mathrm{A}(x_1, \ldots, x_n).$$

Enquanto que a primeira frase afirmava especificamente acerca de cada propriedade que esta ocupa este ou aquele papel e se relaciona desta ou daquela maneira com outras propriedades, a frase de Ramsey afirma que existe uma propriedade ou outra que faz tudo isso. A frase de Ramsey, sendo implicada pela primeira, é uma descrição correta do mundo atual. Agora, imaginemos que se acrescenta à frase de Ramsey a informação de que as propriedades que ocupam os papéis por ela descritos *não* são as mesmas que as que os ocupam atualmente. Temos então esta frase:

$\exists x_1, \ldots, x_n \, A(x_1, \ldots, x_n) \wedge \sim (x_1 = a_1) \wedge \ldots \wedge \sim (x_n = a_n)$.

Podemos construir uma outra frase que nos diz ainda que *nenhuma* das propriedades que ocupam os papeis descritos é uma propriedade exemplificada no mundo atual (i. e., diz-nos que esses papeis são ocupados apenas por propriedades estranhas):

$\exists x_1, \ldots, x_n \, A(x_1, \ldots, x_n) \wedge \sim (x_1 = a_1 \vee \ldots \vee x_1 = a_n) \wedge \ldots \wedge \sim (x_n = a_1 \vee \ldots \vee x_n = a_n)$.

Ao contrário do que acontecia com a frase de Ramsey, estas últimas já não descrevem corretamente o mundo concreto. Mas descrevem-no, ainda que falsamente, com um enorme detalhe e, por isso, o ersatzista poderia dizer que podem descrever completamente como seria o mundo concreto se fosse de uma outra maneira. E, com isto, estariam em condições de identificar cada uma destas frases com mundos ersatz não-atuais (ou com conjunções de frases que seriam os elementos de um mundo ersatz). E, alterando os papeis que são atribuídos às várias propriedades e relações, teríamos descrições de todas as restantes possibilidades.

De acordo com D. Lewis, no entanto, estas suposições são erradas. Se as leis da natureza – e, consequentemente, os papeis nomológicos das várias propriedades – forem contingentes, como D. Lewis pensa que são, então os mundos ersatz construídos a partir das frases de Ramsey são descrições corretas de *várias* maneiras como o mundo poderia ser, e não apenas de uma única. Por isso, não os descrevem completamente: se os descrevessem, seriam capazes de distinguir a possibilidade a que correspondem de todas as outras (a não ser que existissem outras indiscerníveis, uma complicação é aqui irrelevante), mas não o conseguem (Lewis 1986b: 161-5).

Capítulo 2 – A identidade psicofísica

2.1 – Método de definição dos termos teóricos (uma exposição informal)

Em "How to Define Theoretical Terms" (1970c), D. Lewis defende que uma teoria define implicitamente os termos que introduz, e apresenta um método para chegar à definição explícita desses termos. Mais tarde, vai aplicar esse método também aos termos mentais. Por esse motivo, antes de apresentar a teoria da mente de D. Lewis, é conveniente fazer uma exposição informal do método de definição dos termos teóricos e das ideias em que se baseia. (Uma apresentação formal aparece no anexo 1.)

Na terminologia de D. Lewis, considera-se um termo teórico qualquer termo que (1) é introduzido no vocabulário de uma linguagem através do seu aparecimento numa teoria científica e (2) não recebeu qualquer interpretação independentemente do seu uso na teoria em que aparece. É irrelevante se, posteriormente, o termo adquire um uso fora do contexto estritamente científico e passa a fazer parte do vocabulário corrente. Também é irrelevante se nomeia indivíduos ou propriedades (e relações), se nomeia entidades observáveis e familiares ou invisíveis e hipotéticas. E, ainda, é irrelevante se o termo é um predicado, um functor, um nome, ou pertence a qualquer outra categoria sintática. Chamam-se *antigos* aos restantes termos (Lewis 1970c: 428-29).

A teoria que introduz uma classe de termos teóricos afirma dos seus referentes que eles têm estas ou aquelas características e que estabelecem estas ou aquelas relações entre si e com outras coisas. Àquilo que a teoria afirma acerca do referente do termo chamamos o papel teórico associado ao termo. A ideia de D. Lewis é que, não havendo qualquer uso prévio de um termo, ele é implicitamente definido pela teoria em que aparece como nomeando esta ou aquela coisa que ocupa o papel teórico que lhe está associado. Exemplificando:

> «If, without benefit of any prior definition of 'entropy', thermodynamics says that entropy does this, that and the other, we may factor that into two parts. There is an existential claim – a 'Ramsey sentence' – to the effect that there exists some quantity which does this, that and the other (or near enough). And there is a semantic stipulation: let that which does this, that, and the other (or near enough), if such there be, bear the name 'entropy'. Here is another way to say it: the theory associates with the term 'entropy' a certain theoretical role. It claims that this role

> is occupied. And it implicitly defines 'entropy' as a name for the occupant of the role.» (Lewis 1997: 326)

Há, então, uma componente sintética da teoria (expressa pela chamada frase de Ramsey da teoria), que lida com questões de facto, e uma outra que é puramente analítica (expressa pela respetiva frase de Carnap). A teoria afirma que existem coisas que ocupam certos papeis teóricos. Esta é a parte sintética. Mas também afirma, implicitamente, que havendo coisas que ocupam esses papeis, essas coisas são os referentes dos termos teóricos a que os papeis estão associados. Esta é a parte analítica.

Com base nisto, o método para chegar às definições explícitas passa por transformar a conjunção das frases da teoria (o chamado *postulado da teoria*), através de passos simples, em descrições definidas que fornecem, cada uma, a interpretação correta do termo teórico a que correspondem. Seja T a conjunção das frases de uma teoria que introduz o termo *t*. Tudo aquilo que a teoria afirma acerca do referente de *t* está contido em T. Assim, pelo que vimos antes, T contém toda a informação que interpreta *t*. Além disso, *t* ocorre algures em T. Podemos substituir *t* em T por uma variável livre. Então, finalmente, definimos *t* como "o *x* tal que T(*x*)". Temos assim a definição explícita.

Na verdade, o método é mais complexo do que estou a fazer parecer. Assim apresentado simplificadamente, fornece definições circulares quando uma teoria introduz mais do que um termo. Nessa situação, as descrições definidas que definem *t* vão conter os restantes termos introduzidos pela teoria, que, por sua vez, vão ser definidos por descrições que também contêm *t*. O método de D. Lewis resolve este problema através do mecanismo formal das frases de Ramsey (e das frases de Carnap). A estratégia é substituir todos os termos teóricos, à exceção daquele que vai ser definido, por variáveis ligadas por um quantificador existencial, antes de construir a definição definida. Assim, a circularidade viciosa desaparece (*ver* anexo 1).

Surge agora a questão de saber o que acontece à referência de um termo teórico quando a teoria que o introduz não é realizada ou é realizada multiplamente – i. e., quando nada ocupa os papeis teóricos associados aos termos, ou quando esses papeis estão ocupados por mais do que uma sequência de coisas. A resposta de D. Lewis não foi

sempre a mesma entre (Lewis 1970b) e (Lewis 1994). Ele próprio explica aqui essa mudança:

> «I once proposed adding that if a theory has no realization, or multiple realizations, its theoretical terms do not refer. I'd now say that if it is unrealized but almost realized, its theoretical terms refer to the members of its unique near-realization, if there is one; and that if it has multiple realizations (or near-realizations) its theoretical terms have indeterminate reference. My reason for saying that the theoretical terms of multiply realized theory do not refer was that a theorist may be presumed to have intended to implicitly define the terms he introduces. But there is a simpler way to respect the theorist's presumed intention: we should write the postulate in such a way that his theory cannot be multiply realized.» (Lewis 2009: 220, nota 9)

Os termos teóricos introduzidos por uma teoria que nem sequer está perto de ser realizada têm uma referência vazia. (Eles continuam, mesmo assim, a estar definidos pela teoria, tendo um sentido sem um referente no mundo atual.) Mas se a teoria estiver quase realizada, D. Lewis considera que os termos referem as coisas que ocupam imperfeitamente os papeis teóricos a que estão associados. Quando há realização múltipla, D. Lewis admite no seu trabalho mais tardio que estamos perante um caso de indeterminação referencial (Lewis 1994: 417, Lewis 1997: 334), em vez de ausência referencial, como defendera em (Lewis 1970: 432-33).

2.2 – Termos mentais e a psicologia popular

O método de definição dos termos teóricos é crucial para a teoria da mente de D. Lewis, porque ele considera que devemos analisar o vocabulário mental como se fosse verdade o mito de que esse vocabulário foi introduzido na linguagem comum dos nossos antepassados através do seu aparecimento numa teoria criada para explicar e prever o comportamento das pessoas. Essa teoria mitológica corresponde à nossa *psicologia popular* – o agregado de opiniões comuns acerca da nossa vida mental. A ideia de D. Lewis não é que consideremos o mito como verdadeiro, mas que o aceitemos como adequado para alcançarmos uma análise correta dos termos mentais. Como uma hipótese

em história da ciência, o mito é falso. Mas como um procedimento heurístico, pode gerar resultados verdadeiros. É assim que D. Lewis defende esta hipótese:

> «I adopt the working hypothesis that it is a good myth. This hypothesis can be tested, in principle, in whatever way any hypothesis about the conventional meanings of our words can be tested. I have not tested it; but I offer one item of evidence. Many philosophers have found Rylean behaviorism at least plausible; more have found watered down, 'criteriological' behaviorism plausible. There is a strong odor of analyticity about the platitudes of common-sense psychology. The myth explains the odor of analyticity and the plausibility of behaviorism. If the names of mental states are like theoretical terms, they name nothing unless the theory (the cluster of platitudes) is more or less true. Hence it is analytic that *either* pain, etc., do not exist *or* most of our platitudes about them are true. If this *seems* analytic to you, you should accept the myth […].» (Lewis 1972: 257)

O raciocínio de D. Lewis é próximo de uma inferência para a melhor explicação. Temos fortes intuições sobre o significado de termos mentais, que envolvem uma tendência a considerar plausíveis certas variedades de behaviorismo, que associam os estados mentais a certos padrões de comportamento, e a ver como analíticos os princípios da psicologia popular. A hipótese do mito permite explicar estes factos, como ficará evidente pelo que veremos adiante.

Mas o mito é adequado apenas até um certo momento. Há algumas diferenças relevantes entre a psicologia popular e qualquer teoria científica. Em primeiro lugar, contrariamente ao que acontece com a maior parte das teorias científicas, a psicologia popular deve ser conhecimento comum (Lewis 1972: 256). Todos os falantes da linguagem a que pertence o vocabulário psicológico que queremos definir devem conhecer – e considerar verdadeiro – aquilo que é dito por essa teoria, devem saber que os outros falantes também o sabem, devem também saber que os outros falantes sabem que todos os outros o sabem, e por aí em diante. Este é, para D. Lewis, um requisito fundamental para que a psicologia popular forneça o significado dos termos mentais, visto que o significado das expressões que pertencem ao vocabulário de uma linguagem deve ser conhecimento comum entre os seus falantes (*ver* secção 3.5).

Exclui-se da psicologia popular, por isso, aquilo que conhecemos dos estados mentais unicamente através da investigação empírica profissional, e que tem de ser expresso pelo vocabulário de uma teoria científica. A psicologia popular não nos pode

dizer, por exemplo, que certos estados mentais são causados pelo impacto de luz na retina, ou por uma certa atividade neuronal, nem que este ou aquele estado mental causa tais e tais modificações musculares, e coisas desse género. Além de não serem propriedade comum, as teorias empíricas da mente que apareceram na história da ciência são posteriores à existência do vocabulário mental, e por isso não podem servir para analisar esse vocabulário.

Em segundo lugar, o conhecimento da psicologia popular é, como o nosso conhecimento das regras gramaticais, em geral, tácito (Lewis 1994: 416). A maioria de nós é capaz de explicar e prever o comportamento das pessoas através da informação acerca dos seus estados mentais, como é capaz de identificar as frases não muito complexas que são gramaticais. A capacidade para explicar e prever envolve a aplicação de princípios gerais (que podem ser expressos por uma frase universal) a informação sobre factos particulares (Hempel e Oppenheim 1948). Mas poucos de nós são capazes de apresentar as suas crenças sobre os estados mentais de uma maneira explícita e sistemática.

Esta é uma dificuldade clara para qualquer projeto que tenha como objetivo a análise explícita dos termos mentais. Não levanta, no entanto, qualquer dificuldade à proposta geral de D. Lewis em relação ao significado desses termos: eles continuam a ser interpretados como nomes para aquilo que ocupa um certo papel teórico, determinado pelo conhecimento implícito ou explícito que os humanos possuem para explicar e prever o comportamento de outros sujeitos. Isto também não implica, ainda, que o projeto de definição explícita seja impossível. Para o levar a cabo, os hipotéticos investigadores teriam de testar a aceitação de inúmeras frases sobre estados mentais em todos, ou na maioria, dos falantes de uma língua; e, apesar de árdua, esta tarefa não é impossível, pelo menos em princípio.

Mesmo sem ter em mãos princípios explícitos, D. Lewis sugere que estes terão, de um modo geral, a seguinte forma:

> «When someone is in so-and-so combination of mental states and receives sensory stimuli of so-and-so kind, he tends with so-and-so probability to be caused thereby to go into so-and-so mental states and produce so-and-so motor responses.» (Lewis 1972: 256)

Um princípio assim caracteriza um estado mental como estando relacionado causalmente com outros estados mentais, comportamentos e estímulos percetivos. Os estados mentais formam uma rede complexa que serve como mediadora entre os estímulos e o comportamento. Esta rede é afetada pelo impacto de estímulos e, na sua totalidade, leva a certos comportamentos. Podemos dizer que a psicologia popular, com estes princípios, especifica um papel causal para cada estado mental no interior dessa rede.

Assumindo que os termos mentais devem ser interpretados pelo seu aparecimento na psicologia popular, podemos aplicar o método de definição dos termos teóricos (*ver* secção 2.1 e anexo 1). Do referente de cada termo mental 'M', a psicologia popular afirma que R(M), sendo R o papel causal associado a M. Dizemos agora que a descrição definida 'o *x* tal que R(*x*)' fornece o significado de 'M'.

É importante vermos que descrever um estado como ocupando um determinado papel causal não envolve qualquer descrição da natureza desse estado. O vocabulário mental é, de certa forma, ontologicamente neutro:

> «When we describe mental state M as the occupant of the M-role, that is what Smart (1959) calls a topic-neutral description. It says nothing about what sort of state it is that occupies the role. It might be a non-physical or a physical state, and if it is physical it might be a state of neural activity in the brain, or a pattern of current and charges on a silicon chip, or the jangling of an enormous assemblage of beer cans.» (Lewis 1994: 418)

Um estado, físico ou não físico, ocupa um certo papel causal em virtude, em parte, das leis da natureza que prevalecem no mundo em que se encontra e, em parte, do organismo ou sistema em que está integrado. Diferenças nesses aspetos podem fazer com que um estado deixe de ter certas causas e efeitos que atualmente tem. Assim, para qualquer estado E, (1) a propriedade de ser o ocupante de um papel causal é acidental para E, e (2) é possível que outros estados, diferentes de E, ocupem o papel causal de E.

Apesar de ser compatível com teorias que caracterizam os estados mentais como coisas com diferentes naturezas (por exemplo, como propriedades de uma alma simples ou como estados físicos de um organismo complexo), esta análise exige que os estados

mentais sejam causalmente eficazes: «It is not neutral […] between all current theories of mind and body. Epiphenomenalism and parallelist dualism are ruled out as contradictory because they deny the efficacy of experience. Behaviorism as a thoroughgoing dispositional analysis of all mental states, including experiences, is all likewise ruled out as denying the reality and *a fortiori* the efficacy of experiences.» (Lewis 1966: 20) Ou os estados mentais se relacionam causalmente com o mundo físico de uma determinada maneira, ou os estados mentais simplesmente não existem.

2.3 – Algumas circularidades aparentes na análise dos conceitos mentais

Há que ter em conta que o vocabulário que usamos para falar de estímulos e comportamento está, muitas vezes, conceptualmente relacionado com os estados mentais. É um problema para a análise dos termos mentais se a psicologia popular utiliza vocabulário em que isto acontece.

Comecemos por ver o caso dos estímulos. «The causal roles of mental states involve responses to perceptual stimuli» e, D. Lewis lembra-nos, «the relevant feature of the stimulus will often be some secondary quality – for instance, a colour. We cannot replace the secondary quality with a specification of the stimulus in purely physical terms, on pain of going beyond what is known to folk psychology.» (Lewis 1994: 416)

É plausível assumir que a análise correta de 'vermelho' contém, pelo menos em parte, a especificação de que esta expressão nomeia uma propriedade dos objetos que, em condições normais de luminosidade, causam a experiência de vermelho num sujeito. Mas, por sua vez, é parte do papel causal da experiência de vermelho ser causada pela presença de objetos vermelhos. Temos, assim, de excluir a expressão 'vermelho' do vocabulário da psicologia popular, sob pena de ficarmos com uma definição circular de 'experiência de vermelho'.

A solução que D. Lewis apresenta consiste em incluir uma outra teoria – a psicofísica popular – na psicologia popular, e tratar os nomes para cores (sons, odores, e outras propriedades sensíveis) como termos a serem definidos, tal como os termos mentais. Dividimos então a psicofísica popular em duas partes. Uma delas afirma que existem vários pares ordenados que satisfazem o papel funcional complexo descrito pela

frase aberta '*x* é uma propriedade dos objetos que tipicamente causam *y*'. Diz-nos a outra parte que o primeiro elemento do par ordenado é nomeado por 'vermelho', ou 'azul', ou 'verde', ou qualquer outro nome para uma cor, conforme o caso, e que o segundo elemento é nomeado por 'experiência de vermelho', ou 'experiência de azul', ou 'experiência de verde', ou qualquer outro nome para a experiência de uma cor.[27] Neste método não há qualquer circularidade na definição, assim como não há circularidade ao definir todos os termos mentais – ou os termos teóricos de uma qualquer teoria – de uma só vez (*ver* anexo 1).

Do outro lado do problema está a descrição do comportamento. Queremos considerar como comportamento apenas os meros movimentos do corpo de um sujeito, mas normalmente falamos destes através de descrições psicologicamente carregadas. Creio que estas descrições podem ser divididas pelo menos em três tipos. (1) Existem descrições de comportamento que implicam a atribuição de certos estados mentais a um

---

27 É importante ter em conta que o papel causal complexo atribuído às cores e às experiências dessas cores não é descrito totalmente pela frase aberta '*x* é uma propriedade dos objetos que tipicamente causam *y*'. Qualquer par ordenado formado por uma cor e a respetiva experiência satisfaz esta frase. Se as definições dos termos dependessem de uma teoria assim, a sua denotação seria radicalmente indeterminada (Lewis 1997: 335).

Uma solução apresentada por D. Lewis consiste em acrescentar informação relativa às várias cores na psicofísica popular. Se conseguirmos determinar um dos elementos dos vários pares ordenados, temos a esperança de que o outro possa também ser determinado. Neste caso, se tivermos a cor teremos também a experiência correspondente.

No entanto, como D. Lewis nota, isto é feito através de informação que não é conhecimento comum entre a comunidade de falantes de uma certa linguagem. E, se pretendemos analisar os termos para cores e experiências através da nossa teoria psicofísica, a situação torna-se problemática, já que o significado dos termos de uma linguagem são conhecimento comum entre os falantes dessa linguagem, ou pelo menos entre todos os que usam tais termos (*ver* secção 3.5). Essa informação consistiria, por exemplo, em dizer que o amarelo é a cor desta e daquela coisa com a qual já estivemos em contacto, ou que o azul é a cor à qual daríamos a resposta 'azul' se nos perguntassem, apontando para um objeto dessa cor, 'Que cor é esta?', ou, ainda, que o cinzento é a cor que tipicamente causa a experiência que estou agora a ter ou a imaginar, e por aí em diante (Lewis 1997: 335-37).

D. Lewis resolve este problema distinguindo o conhecimento comum da definição de um termo de um tipo de conhecimento comum *existencial*. Se tentarmos escrever uma definição para cada uma das cores a partir das várias informações que temos sobre as mesmas, conseguiremos no máximo formular uma definição que é conhecimento comum numa parte – na pior, mas mais provável das hipóteses, uma parte composta por apenas uma pessoa – da comunidade de falantes. Neste sentido, não há conhecimento comum da definição. Mas, para haver comunicação, pode ser suficiente que seja conhecimento comum que existe uma definição dos termos para cada parte da comunidade de falantes e que haja conhecimento comum de que as várias definições do mesmo termo são satisfeitas pelas mesmas coisas. Em casos deste género, para evitar confusões comunicacionais – principalmente em contextos modais e contrafactuais –, D. Lewis considera ainda relevante que as várias definições existentes na comunidade de falantes sejam rigidificadas, i. e., que nos digam que vermelho é a cor desta e daquela coisa em @, aqui, e agora (Lewis 1997: 339-342).

agente: ‘X está a fugir de Y’ implica que, além de X estar a movimentar-se de uma certa maneira, X saiba que Y se encontra por perto, que X deseje manter Y fora do seu alcance, e que X acredite que levará a cabo o seu desejo se se movimentar desta ou daquela maneira. Em *Intention* (1957), Elizabeth Anscombe fornece uma lista de verbos que são usados em descrições que implicam uma certa intenção por parte do agente: ‘telephoning, calling, groping, crouching, signing, signalling, paying, selling, buying, hiring, dismissing, sending for, marrying, contacting’ (Anscombe 1957: 84-5). (2) Outras descrições não implicam a atribuição de estados mentais, mas envolvem conceitos que são definidos a partir de conceitos mentais. De ‘X disse que P*’* (citação indireta), por exemplo, não se segue que X tenha este ou aquele estado mental (X pode dizer que P estando adormecido), mas o predicado ‘disse que P’ implica que há um padrão de estados mentais normalmente envolvidos na comunicação entre membros da comunidade de X. Mais uma vez, G. E. M. Anscombe fez o inventário de alguns verbos que envolvem o conceito de intenção: ‘intruding, offending, coming to possess, kicking (and other descriptions connoting characteristically animal movement), abandoning, leaving alone, dropping (transitive), holding, picking up, switching (on, off), placing, arranging’ (Anscombe 1957: 84-5). (3) Finalmente, outras descrições ainda pressupõem algumas características (sociais e culturais) salientes das circunstâncias em que o comportamento tem lugar, que não podem ser compreendidas sem o recurso à psicologia popular. D. Lewis o exemplo de ‘X pontapeou a bola para o seu colega de equipa’ (Lewis 1994: 417).

Os princípios da psicologia popular não podem estar limitados a falar dos vários padrões de comportamento unicamente através de descrições puramente físicas, mesmo que baseadas numa teoria física popular, como, por exemplo, ‘X está a correr’, ‘X emitiu a frase “tenho frio”’ e ‘X está a pontapear uma bola’. O problema é que estes princípios, por um lado, não são válidos para muitos movimentos tão estreitamente descritos e, por outro, aplicam-se a classes mais amplas de movimentos que deixam assim de ser referidos na psicologia popular. Suponhamos que parte do papel causal da dor é fazer com que o sujeito fuja daquilo que está a provocar um certo dano causador da dor. Como é que podemos descrever este comportamento? Se dissermos que este consiste em fugir para longe do perigo, então a psicologia popular está severamente enganada e incompleta.

Correr não é a única maneira de fugir. Por vezes, é conveniente escapar a um perigo com movimentos lentos, ou até permanecendo parado. Suponhamos também que parte do papel causal da crença de que P é que esta leve um sujeito, dadas outras crenças e certos desejos, a dizer frases que transmitem a informação de que P é verdadeira. Dizer frases que transmitem esta informação pode ser emitir 'Estou com frio' ou 'Não!', dependendo das convenções linguísticas prevalecentes.

A proposta de D. Lewis é que se descreva o comportamento através do efeito que este tem em certas circunstâncias. Dizemos que X movimenta-se de um tal modo que, *se* X estiver nas circunstâncias tais e tais, *então* o seu movimento terá este ou aquele impacto (Lewis 1994: 417). Por exemplo, 'X está a fugir de Y' deve ser trocada por 'X está a fazer movimentos tais que, nas circunstâncias em que X se encontra, levam a que X se mantenha afastado de Y'. 'X está a dizer que P' pode ser substituída por 'X está a dizer uma frase que, nas circunstâncias em que X se encontra, o seu ouvinte vai provavelmente passar a acreditar que X acredita que P e, se confiar em X, vai acreditar que P'. Esta descrição envolve conceitos psicologicamente carregados que, creio eu, não podem ser eliminados. Mas provavelmente o caso também não é grave: se o papel causal das crenças não depender apenas do comportamento linguístico dos falantes (*ver* capítulo 3), o que acontece é que, em vez do comportamento descrito normalmente por 'X está a dizer que P' contribuir para a definição funcional dos termos mentais, é a psicologia popular que permitirá definir o que está envolvido nessa descrição comportamental.

Descrever assim o comportamento permite escapar ao *chauvinismo*, como é notado em (Braddon-Mitchell e Jackson 2007: 60). Uma teoria da mente é chauvinista se tiver como consequência a negação de vida mental a sujeitos que intuitivamente a possuem, por restringir a atribuição de estados mentais a indivíduos com características físicas, químicas, ou neurológicas, específicas. Envolvendo nos princípios da psicologia popular descrições da física popular, como 'X está a movimentar o braço rapidamente', estamos a excluir a aplicação dos termos mentais a sujeitos sem braços (Block 1980: 294-95).[28]

[28] É forte, no entanto, o argumento de que uma teoria da mente com os contornos da de D. Lewis é chauvinista relativamente a casos como (1) o dos cérebros numa cuba e (2) dos paralíticos, como nota Ned Block em (1980: 283-84). Repare-se que em (1), mesmo existindo relações causais entre o meio circundante

Há ainda uma aparente circularidade nas descrições dos estímulos e comportamentos utilizadas nas generalizações da psicologia popular acerca das *atitudes intencionais* – aqueles estados mentais, como crenças e desejos, que têm um certo conteúdo representacional. Contrariamente ao que acontece com a experiência de vermelho ou a dor, não existe uma classe específica de estímulos e comportamentos que estão associados em geral a cada uma das atitudes. É apenas quando um conteúdo – uma proposição, por exemplo[29] – é indexado às atitudes que estes estados mentais passam a ter uma relação mais ou menos direta com certos estímulos e comportamentos. Vestir um casaco é um comportamento normalmente causado pela crença de que está a ficar frio, mas muito dificilmente pela crença de que está a ficar calor, pelo menos numa pessoa com uma motivação comum.

Tendo em conta o número bastante grande – e talvez infinito – de proposições que as atitudes podem ter como conteúdo, é impossível que a psicologia popular contenha uma generalização a associar uma classe específica de estímulos e comportamentos a cada par de atitudes e proposições. Em vez disso, a psicologia popular vai precisar de tratar as atitudes como estados relacionais entre um sujeito e uma proposição, e ter um número finito de generalizações que especifiquem a relação geral entre os vários conteúdos possíveis e os estímulos e comportamentos a que as atitudes com esses conteúdos vão estar associadas.[30] Essas generalizações terão de ter, mais ou menos, uma destas formas, sendo E e C, respetivamente, letras esquemáticas para frases sobre estímulos e comportamentos:

---

e o interior do cérebro, estas não são plausivelmente aquelas que estão especificadas pela psicologia popular; e em (2) um indivíduo nessas condições pode deixar de ter um grande número de possibilidades de afetar o ambiente próximo pelo seu comportamento, deixando os seus estados mentais de desempenhar partes fundamentais do seu papel causal. Parte destes casos podem ser resolvidos pela admissão de D. Lewis de que os estados mentais não têm de ocupar em todos os casos, sem exceção, o papel causal característico, como se verá adiante, na secção 2.5.

29 Aqui vou falar como se as proposições fossem os únicos conteúdos adequados às atitudes. No entanto, mais à frente, no capítulo 4, veremos que D. Lewis considera que, para além de proposições, vamos precisar de considerar propriedades como o conteúdo de algumas atitudes – mais concretamente, das atitudes irredutivelmente egocêntricas.

30 Esta é a proposta de D. Lewis em (1972: 256, nota 13). Devemos ter em conta que se a psicologia popular, não podendo conter generalizações sobre cada uma das atitudes em particular, não as tratar como estados relacionais, fica ameaçada a análise de todos os nomes para estados mentais.

(1) Para todo o sujeito S e toda a proposição P, se E e ..., então S vai estar na atitude A com conteúdo P.

(2) Para todo o sujeito S e toda a proposição P, se S está na atitude A com conteúdo P e ..., então C.

A ameaça de circularidade aparece quando se começa a pensar que frases podem aparecer no lugar das letras esquemáticas E e C. Essas frases não podem descrever uma classe específica de estímulos e comportamentos, porque a adequação de uma classe desse género a uma atitude é variável com o conteúdo. Teremos então de dizer que os estímulos e comportamento adequados à atitude A com conteúdo P são aqueles que estão normalmente associados à atitude A com conteúdo P? Mas isso é obviamente circular, porque na descrição dos estímulos e comportamentos vão aparecer os conceitos das atitudes. A solução passa por encontrar descrições que excluam esses conceitos, mantendo, no entanto, o conceito de proposição. Há algumas descrições suficientemente abstratas que permitem especificar uma classe de estímulos e comportamentos a partir das proposições, como por exemplo 'X é confrontado com evidência de que P', 'X comporta-se de modo a tornar verdade que P', 'X disse uma frase que expressa P'. Reformulamos, então, (1) e (2) deste modo:

(1*) Para todo o sujeito X e toda a proposição P, se E(P) e ..., então X está numa atitude com conteúdo P.

(2*) Para todo o sujeito X e toda a proposição P, se X está numa atitude com conteúdo P e …, então C(P).[31]

[31] Alguém poderia agora levantar a objeção de que continuamos a precisar de infinitos nomes, não para estados mentais, mas para as proposições que os estados mentais vão relacionar a um sujeito. Mas esta objeção falha pelo facto de não precisarmos de nomes para qualquer objeto que faça parte do domínio de quantificação das variáveis que usamos numa teoria. E o único que necessitamos para descrever o papel das crenças e desejos em geral é de quantificar sobre todas as proposições, e não de nomeá-las.

2.4 – Da análise conceptual à identidade psicofísica

A teoria da identidade psicofísica desdobra-se em duas variedades de diferente força. Na variedade mais fraca identificam-se os espécimes (*tokens*) dos vários estados mentais com espécimes de certos estados físicos. Na variedade mais forte são os próprios tipos de estados mentais que são identificados com certos tipos de estados físicos. Um espécime é uma ocorrência particular numa certa região espaciotemporal, enquanto que um tipo é repetível em várias ocorrências. Distinguimos tipos e espécimes quando dizemos, por exemplo, que X e Y estão a ter a *mesma* experiência, estão a ver um objeto vermelho, mas *a experiência de* X começou há mais tempo que a *de* Y. Desde "An Argument for the Identity Theory" (Lewis 1966), D. Lewis defende a variedade mais forte desta teoria.

(Daqui em diante, falo em *estados* (mentais ou físicos) para me referir a tipos, utilizando maiúsculas ('M' e 'F') como letras esquemáticas para os nomes desses tipos. As minúsculas ('m' e 'f') serão usadas para espécimes.)

O argumento apresentado em (Lewis 1966, 1972, 1994) envolve a análise funcional que estivemos a considerar. Recapitulando, esta análise afirma que, para qualquer estado mental M, é verdade que, em virtude do significado de 'M':

(1) M = o ocupante do papel causal R.

Como vimos antes, (1) é compatível tanto com uma ontologia materialista como dualista da mente. Para restringir as possibilidades que plausivelmente podem ser atuais, D. Lewis propõe que aceitemos certas hipóteses empiricamente motivadas. Estas hipóteses têm como consequência que, para qualquer papel causal R associado pela psicologia popular a um estado mental, sendo F um estado físico qualquer:

(2) O ocupante do papel causal R = F.

Podemos dizer que (1) é a componente analítica do argumento, e (2) é a componente *a posteriori*. Em (Lewis 1994), (2) é defendida com o recurso a uma hipótese

de *superveniência* materialista. De um modo pouco rigoroso, a hipótese pode ser apresentada como dizendo que o carácter qualitativo do mundo não pode variar sem uma diferença a nível do padrão de instanciação das propriedades físicas fundamentais pelos vários pontos espaciotemporais (Lewis 1994: 412-13).[32] E, como D. Lewis afirma: «if materialist supervenience is true, and every feature of the world supervenes upon fundamental physics, then the occupant of the role is some physical state or other – because there's nothing else for it to be.» (Lewis 1994: 418) (No anexo 2 apresento, resumidamente, a resposta de D. Lewis ao *argumento de conhecimento* contra a superveniência materialista.)

Outra abordagem foi seguida antes, em (Lewis 1966), apresentada assim:

> «My second premise is the plausible hypothesis that there is some unified body of scientific theories, of the sort we now accept, which together provide a true and exhaustive account of all physical phenomena (i.e. all phenomena describable in physical terms). They are unified in that they are cumulative: the theory governing any physical phenomenon is explained by theories governing phenomena out of which that phenomenon is composed and by the way it is composed out of them. The same is true of the latter phenomena, and so on down to fundamental particles or fields governed by a few simple laws, more or less as conceived of in present-day theoretical physics.» (Lewis 1966: 23)

A ideia geral é a seguinte. Qualquer tipo de fenómeno físico, η_1, η_2, …, pode ser tratado por uma teoria T_1, a qual é explicável por (ou redutível a) uma teoria T_2 relativa aos fenómenos físicos, η_3, η_4, …, que são os constituintes de η_1, η_2, …, juntamente com a descrição de como η_3, η_4, … se relacionam para formar η_1, η_2, …. O que acontece com T_1 relativamente a T_2 aplica-se a T_2 relativamente a uma outra teoria T_3, ainda mais fundamental. A certo ponto, uma teoria T_n, que explica T_1, T_2, … será explicada por T_F, a teoria que trata dos constituintes mais básicos da realidade física. Deste modo, T_F explica também T_1, T_2, …, T_n, juntamente com a descrição de como as partículas ou os

[32] É parte desta tese a defesa de que as únicas propriedades perfeitamente naturais (*ver* secção 3.4) instanciadas no mundo possível atual são propriedades físicas (não necessariamente aquelas que aparecem nas teorias físicas presentes, mas as que serão referidas numa hipotética teoria definitiva e última, a qual D. Lewis acredita não estar muito distante do conhecimento presente). D. Lewis defende que quaisquer mundos possíveis idênticos no padrão de instanciação de propriedades perfeitamente naturais são idênticos *simpliciter* (Lewis 1992a: 218-19). Daqui se segue que qualquer mundo idêntico ao nosso no que diz respeito à instanciação de propriedades físicas, e no qual não sejam instanciadas propriedades perfeitamente naturais *estranhas* ao mundo atual, será idêntico a este (Lewis 1994: 412-13).

campos – ou qualquer outro tipo de entidades – tratados em T_F se relacionam para formar fenómenos complexos como η_1, η_2, η_3, η_4, ….

Contrariamente à hipótese de superveniência materialista, este princípio não é uma tese ontológica – ou seja, não nega a existência de qualquer fenómeno não físico. Não é, também, uma negação da interação entre a parte física do mundo e uma possível parte não física. Não só podem existir fenómenos não físicos, como estes podem ser causados por fenómenos físicos, e esta relação pode até apresentar regularidades descritíveis como leis da natureza. No entanto – e isto é o que é relevante para a questão que aqui nos prende – os fenómenos não físicos devem ser ineficazes em relação aos fenómenos físicos. A ocorrência dos primeiros não tem, segundo este princípio, qualquer impacto nos últimos. Se assumirmos, como parece plausível, que um fenómeno físico é um qualquer fenómeno tratado pelas teorias T_1, T_2, …, T_n, … então, segundo o princípio de D. Lewis, qualquer fenómeno físico é explicável recorrendo apenas a princípios que governam fenómenos físicos. Em última análise, estes princípios são redutíveis às leis apresentadas em T_F, com o auxílio das descrições de como aquilo que é tratado por T_F se relaciona de modo a formar o fenómeno que se pretende explicar.

Visto que o papel causal da maior parte dos estados mentais envolve o comportamento como efeito, e tendo em conta que o comportamento é um fenómeno físico, então se o papel causal for realizado, o seu ocupante terá de ser um estado físico (Lewis 1966: 24). Em outros casos, em que o papel causal de um estado mental não envolve comportamento, mas tem como efeito apenas outros estados mentais, não é possível afirmar diretamente alguma coisa sobre o seu ocupante, a partir deste princípio. No entanto, é muito provável que, mesmo não envolvendo o comportamento como efeito, o papel causal de um estado mental envolva sempre pelo menos alguns estados mentais que causam comportamento. Como estes últimos têm de ser estados físicos, o ocupante do papel causal que os envolve como efeitos tem também de ser um estado físico.

Chegamos, assim, à conclusão do argumento. Dado que, para cada M e alguns R e F, M = o ocupante do papel causal R e o ocupante do papel causal R = F, por

transitividade da identidade conclui-se que M = F.[33] Os estados mentais são idênticos a estados físicos – provavelmente, diz-nos D. Lewis, padrões de atividade neuronal (Lewis 1966: 24, 1972: 249). A partir da análise funcional dos termos mentais e da suficiência explicativa da física em relação aos fenómenos físicos, fica assim estabelecida a teoria da identidade psicofísica.[34]

---

33 Esta é a aplicação de um esquema geral de redução de uma teoria T_1 a uma teoria T_2, ao caso particular da psicologia popular e de uma qualquer hipotética teoria fisiológica que aborde a realização física dos estados mentais. Este esquema segue-se do método de definição dos termos teóricos (*ver* secção 2.1 e anexo 1).

T_1 contém os termos teóricos $t_1, \ldots, t_n$, e T_2 contém entre os seus termos $\alpha_1, \ldots, \alpha_n$ (não é relevante se estes são termos teóricos ou termos antigos). T_1 afirma que $T(t_1, \ldots, t_n)$ e T_2 afirma, entre outras coisas, que $T(\alpha_1, \ldots, \alpha_n)$, e que apenas as entidades nomeadas por $\alpha_1, \ldots, \alpha_n$ satisfazem a frase de realização $T(x_1, \ldots, x_n)$, que é, por sinal, a frase de realização de T_1.

Com o método de definição dos termos teóricos de D. Lewis podemos obter as seguintes leis-ponte para cada termo teórico de T_1 (e, assim, reduzir T_1 a T_2):

$\alpha_1 = t_1$.
…
$\alpha_n = t_n$.

Isto acontece porque de T_2 se seguem estas afirmações de identidade:

$\alpha_1 = o\ x_1 : \exists_1 x_2, \ldots, \exists_1 x_n\ T(x_1, x_2 \ldots, x_n)$.
…
$\alpha_n = o\ x_n : \exists_1 x_1, \ldots, \exists_1 x_{n-1}\ T(x_1, \ldots, x_{n-1}, x_n)$.

E estas últimas são equivalentes às leis-ponte, visto que cada um dos termos $t_1, \ldots, t_n$ é definido por uma das descrições que se encontram nestas afirmações de identidade. Temos assim uma redução de T_1 a T_2 implicada apenas por T_2 (Lewis 1970b: 441-45).

Se assumirmos, no entanto, que T_2 não exclui a sua realização múltipla, não podemos dizer que as leis-ponte se sigam dessa teoria isoladamente. Mas, visto que de T_1 se seguem definições explícitas de $t_1, \ldots, t_n$, de acordo com as quais cada um destes termos nomeia um dos elementos da única sequência que satisfaz $T(x_1, \ldots, x_n)$, se T_2 afirma que essa frase é realizada pela sequência $(\alpha_1, \ldots, \alpha_n)$, então daí se segue que os objetos desta sequência são os referentes de $t_1, \ldots, t_n$. Assim, podemos dizer que as mesmas leis-ponte são deriváveis de T_1 e T_2 – mesmo quando T_2 é enfraquecida.

34 A psicologia popular não especifica apenas um papel causal para cada estado mental, mas também para várias propriedades desses estados. Exemplos disso são os *qualia* – as propriedades que dizem respeito ao aspeto fenomenal de uma experiência, a como é estar a ter essa experiência. A identificação de M com F implica a identificação do quale de M e uma propriedade (física) de F.

Este resultado é problemático, como veremos, apesar de ser a única alternativa para um materialista que seja realista relativamente aos qualia. De acordo com D. Lewis, faz parte do conceito dos qualia que, quando temos uma experiência com quale Q, somos capazes de identificar que Q é uma propriedade dessa experiência (Lewis 1995: 141). E, acrescenta ainda, «the knowledge I gain by having an experience with quale Q enables me to know what Q is – identifies Q – in this sense: any possibility not ruled out by the content of my knowledge is some in which it is Q, and not any other property instead, that is the quale of my experience. Equivalently, when I have an experience with quale Q, the knowledge I thereby gain reveals the essence of Q: a property of Q such that, necessarily, Q has it and nothing else does» (Lewis 1995: 142).

Compare-se esta situação com o seguinte caso. X acredita que a água é imprescindível à vida, mas X não faz ideia de que a água é H2O. Através da sua crença, X divide as possibilidades em duas classes: a

2.5 – Realização múltipla, contingência, rigidez e os casos da dor marciana e da dor louca

Identificar os estados mentais com certos estados físicos levanta alguns problemas (Putnam 1980: 228). (Na apresentação dos problemas, vou assumir que a análise proposta por D. Lewis está correta, apesar de poderem ser apresentados de uma maneira mais neutra.) Um deles é o problema da *realização múltipla*. Este é o problema dos marcianos, discutido por D. Lewis em "Mad Pain and Martian Pain" (1980b). Tanto os humanos como os marcianos sentem dor. Isso acontece porque ambos estão num estado que, entre

daqueles mundos que estão de acordo com aquilo em que acredita, e a dos restantes. Mas qual é a classe em que ele acredita estar o mundo atual? A de todos os mundos em que a *água* é imprescindível à vida? Talvez não, visto que esta classe exclui mundos em que XYZ ocupa o papel atribuído popularmente à água, e X não consegue excluir alguns desses mundos de serem o atual. Conseguiria isso, sim, se acreditasse que a água era H2O. Mas, até não ter essa informação em mente, esta é uma classe inadequada para capturar o conteúdo da sua crença. A resposta mais certa talvez seja a de que a classe de mundos adequados à sua crença é aquela que contém mundos em que existe um líquido que corresponde à nossa conceção popular de água e em que esse líquido é imperscindível. Ora, dizer que quando temos uma experiência com quale Q, passamos a saber o que Q é essencialmente, consiste em dizer que temos um conhecimento de Q como aquele que X teria se descobrisse que a água é H2O: seríamos capazes de restringir as possibilidades admitidas pelas nossas crenças a apenas aquelas que contêm Q – e não algum substituto de Q subjetivamente indiscernível.

Esta ideia não esgota o conteúdo do conceito de qualia. No entanto, é suficiente para levar D. Lewis a dizer que qualquer materialista deve rejeitar a existência de propriedades que cumprem o papel aqui atribuído aos qualia. Se a hipótese materialista estiver correta, um quale de uma experiência é uma propriedade física complexa – talvez a propriedade de ser uma estrutura neuronal, formada deste ou daquele modo a partir dos constituintes físicos básicos. Ora, de acordo com o que a psicologia popular afirma sobre os qualia, quando temos uma experiência, devemos identificar precisamente essas propriedades neuronais – as possibilidades admitidas são apenas aquelas em que a experiência tem essas propriedades. Mas, como afirma D. Lewis, «If qualia are physical properties of experiences, then it is certain that we seldom, if ever, identify the qualia of our experiences», sendo a razão para isso a de que «making discoveries in neurophysiology is not so easy!» (Lewis 1995: 142) (Como D. Lewis nota, se identificarmos os qualia, eles não podem ter uma estrutura da qual permanecemos ignorantes mesmo quanto temos uma experiência com um certo quale. Mas isto é exatamente o que acontece, já que se os qualia são propriedades neuronais, então são propriedades estruturais – propriedades de objetos cujos constituintes estão relacionados de um determinado modo – e essa estrutura não nos é revelada pela experiência (Lewis 1995: 142, nota 4).)

Ao admitir uma posição materialista, chega-se assim à rejeição de que o papel dos qualia é totalmente realizado. Para alguns, esta pode ser uma razão para abandonar o materialismo, enquanto que para outros estes resultados dizem-nos que talvez tenhamos de dizer que os qualia não existem – nenhuma experiência tem essas propriedades (esta seria uma posição eliminativista em relação aos qualia). D. Lewis tende a concordar com esta última posição, mas acredita que ainda se pode falar em qualia, se forem entendidos num sentido mais fraco. A psicologia popular diz-nos também que os qualia são aquelas propriedades responsáveis pela aquisição de habilidades por parte de um sujeito que tem uma certa experiência pela primeira vez – aquelas habilidades que constituem aquilo que é conhecer *como é* ter uma certa experiência (*ver* anexo 2). Existem propriedades materialisticamente aceitáveis que desempenham este papel e podem receber o nome de qualia, ainda que sejam ocupantes imperfeitos do papel causal envolvido no conceito de qualia (Lewis 1995: 141-43).

outras coisas, é causado por danos corporais e causa o desejo de evitar o que está a provocar esse estado – em suma, em ambos, algum estado ocupa o papel causal que a psicologia popular atribui à dor.

No entanto, os humanos e os marcianos são fisicamente muito diferentes. Essa diferença é notória também nos ocupantes do papel causal da dor. Enquanto que nos humanos esse papel é ocupado pelo disparo de fibras-C, nos marcianos é ocupado por uma atividade complexa do sistema hidráulico responsável pelas suas atividades mentais (Lewis 1980b: 216). O problema que este exemplo coloca é, então, o seguinte. Se seguirmos a teoria da identidade e considerarmos que a dor é idêntica ao disparo de fibras-C, exclui-se o marciano de sentir dor. O mesmo se verifica inversamente, se considerarmos – como provavelmente considerariam os marcianos! – que a dor é idêntica a um padrão de atividade hidráulica. (Excluimos, obviamente, a hipótese de dizer que o disparo de fibras-C é idêntico ao padrão de atividade hidráulica.)

O outro problema é que temos de aceitar que qualquer objeto, atual ou meramente possível, que está em F está também em M, se aceitarmos que M é idêntico a F. De um modo equivalente, é impossível que qualquer objeto, atual ou meramente possível, não esteja em M, mas esteja em F. No entanto, F pode desempenhar papeis causais muito distintos em diversos contextos – pode variar entre mundos possíveis ou entre organismos pertencentes ao mesmo mundo. O estado F pode não ocupar num indivíduo X o papel causal que a psicologia atribui a M e, mesmo assim, estamos obrigados a dizer que X está em M, quando está em F.

Estes problemas são dois lados da mesma moeda. Em ambos os casos, temos em mãos uma consequência inaceitável da impossibilidade de variação independente entre estados mentais e os estados físicos com que são identificados.

A teoria da identidade entre espécimes não tem estas consequências. Se, em vez de identificarmos M com F, identificarmos certas instâncias de M (m_1, m_2, …) com algumas instâncias de F (f_1, f_2, …), e permanecermos em silêncio acerca das outras ocorrências destes estados, podemos acomodar o caso dos marcianos e o caso das criaturas sem mentalidade. D. Lewis, no entanto, não segue este caminho. Comentando o

argumento de Hilary Putnam contra a teoria da identidade (baseado no problema da realização múltipla), apresenta assim, em resumo, a sua proposta:

> «Putnam argues that the brain-state hypothesis (and with it, the functionally-specified-brain-state-hypothesis) ought to be rejected as scientifically implausible. He imagines the brain-state theorist to claim that all organisms in pain – be they men, mollusks, Martians, machines, or what have you – are in some single common nondisjunctive physical-chemical brain state. Given the diversity of organisms, that claim *is* incredible. But the brain-state theorist who makes it is a straw man. A reasonable brain-state theorist would anticipate that pain might well be one brain state in the case of men, and other brain (or nonbrain) state in the case of mollusks. […] no one says that the *concept* of pain is different in the case of different organisms […]. But it is the *fixed* concept expressed by 'pain' that determines how the denotation of 'pain' varies with the nature of the organism in question.» (Lewis 1969b: 25)

A solução que D. Lewis apresenta para estes problemas é, então, a de dizer que os nomes para estados mentais não são designadores rígidos – têm uma flexibilidade referencial que é determinada, ainda assim, por um conceito fixo em todos os casos (Lewis 1980b: 218-19, 1994: 418-19).

Como é explicado em (Kripke 1980: 53-60), há, em geral, uma certa ambiguidade quando se está a definir um termo através de uma afirmação de identidade que envolve uma descrição definida. Há dois papeis que a descrição pode estar a desempenhar na definição: pode estar (1) a *fixar* a referência ou (2) a fornecer o significado do termo. Suponhamos que queremos definir o nome próprio 'Aristóteles'. Alguém pode propor como definição a afirmação (verdadeira) de identidade 'Aristóteles = o professor de Alexandre Magno'. Esta afirmação diz-nos que 'Aristóteles' refere, no mundo atual, o professor de Alexandre Magno. Mas há diferenças significativas naquilo que a afirmação diz sobre a referência de 'Aristóteles' nos restantes mundos, se entendermos a descrição 'o professor de Alexandre Magno' como estando a fixar a referência ou a fornecer o significado de 'Aristóteles'.

Se estiver a fixar a referência, 'Aristóteles' vai referir rigidamente – i. e., em todos os mundos – aquilo que, no mundo em que ocorre a definição (para nós seria o mundo atual), é o professor de Alexandre Magno. (A intensão de 'Aristóteles' é, assim, uma função constante.) Pelo contrário, se estiver a fornecer o significado, 'Aristóteles' vai

referir, em cada mundo, aquilo que nesse mundo é o professor de Alexandre Magno. Acontecendo isto, 'Aristóteles' e 'o professor de Alexandre Magno' são sinónimos, e não têm diferentes propriedades semânticas. Noutros mundos possíveis, certamente pessoas diferentes de Aristóteles são o professor de Alexandre Magno. Num desses mundos, é Platão que satisfaz essa descrição. Nesse mundo, 'Aristóteles' refere 'Platão'. Quando a referência de um nome é fixada por uma descrição, a referência das duas expressões apenas tem de ser idêntica no mundo em que a definição ocorre. Mas quando a descrição fornece o significado, a referência do nome acompanha a referência da descrição. Neste cenário, quando a descrição é um designador não-rígido, o nome também o é.

A estratégia de D. Lewis é conceber as descrições funcionais como fornecendo o significado dos termos mentais, como defende para os termos teóricos em geral (Lewis 1970: 435-36). Devemos ter em conta que se a descrição funcional de 'M' fixasse a referência desse termo, a afirmação de que M ocupa este ou aquele papel causal passaria a ser uma verdade relativa a uma questão de facto contingente, em vez de uma verdade analítica necessária, como D. Lewis defende que acontece. Mantendo 'M' e a corresponde descrição como referencialmente coincidentes ao longo dos mundos, não vamos encontrar nenhum mundo em que M não ocupa o papel causal descrito.[35]

Como a descrição funcional fornece o significado sem fixar a referência, diferentes estados em vários mundos podem ser o referente de 'M'. Temos, por isso, uma identidade contingente entre M e F (Lewis 1980b: 218, 1994: 418). Admitir isto não implica que «it might have been that pain was not pain and nonpain was pain» (Lewis 1980b: 218).[36] E, ainda, «I do not say that here we have two states, pain and some neural

---

35 Estamos a lidar, neste caso, com uma necessidade *de dicto* e não *de re*. (Uma necessidade meramente verbal, como afirma D. Lewis em (Lewis 1994: 417).) É verdade, de acordo com esta perspetiva, que:

(1) $\Box \exists x$ (x = M se e só se x = o ocupante do papel causal R).

Mas não é verdade, no entanto. que:

(2) $\exists x$ (x = M se e só se $\Box$ (x = o ocupante do papel causal R)).

36 A situação problemática seria aquela em que era verdade que:

(1) $\Diamond \exists x \sim(x = M) \& (x = M)$.

Mas a contingência admitida por D. Lewis permite-nos apenas dizer algo como:

(2) $\exists x$ (x = M) & $\Diamond \sim (x = M)$.

Ou seja, não existe qualquer mundo – não é possível, portanto – em que aquilo que é M nesse mundo não seja M (nesse mesmo mundo). Mas é possível que aquilo que é M no mundo atual não seja M noutro mundo possível.

state that are contingently identical, identical at this world but different at another. Since I'm serious about the identity we have not two states but one. This one state, this neural state which is pain, is not contingently identical to itself. It does not differ from itself at any world. Nothing does.» (Lewis 1980b: 218)

A escolha entre interpretar a afirmação de identidade entre um estado mental e o ocupante de um papel causal como uma definição propriamente dita ou uma fixação de referência não é óbvia. A ambiguidade que se verifica em geral é ainda agravada em contextos da linguagem comum, e por isso parece não haver evidência decisiva para optar por uma das alternativas. Para D. Lewis, no entanto, há boas razões para fazer esta escolha, como a que apresenta neste argumento:

> «Here is an argument for that 'pain' is not a rigid designator. Think of some occasion when you were in severe pain, unmistakable and unignorable. All will agree […] that there is a state that that actually occupies the pain role (or near enough); that it is called 'pain'; and that you were in it on that occasion. For now, I assume nothing about the nature of this state, or about how it deserves its name. Now consider an unactualized situation in which it is some different state that occupies the pain role; and in which you were in that different state; and which is otherwise as much like the actual situation as possible. Can you distinguish the actual situation from this unactualized alternative? I say no, or not without laborious investigation. But if 'pain' is a rigid designator, then the alternative situation is one in which you were not in pain, so you could distinguish the two very easily. So 'pain' is not a rigid designator.» (Lewis 1994: 419)

.

O caso que D. Lewis está a apresentar é um alegado exemplo de realização múltipla, no qual se considera um único indivíduo X em duas situações possíveis – uma atual, em @, e outra contrafactual, num mundo possível *w*, (1) idêntico a @ em todos os aspetos à exceção do ocupante do papel causal de dor em X. Assume-se também que (2) em @, X está inequivocamente a sentir dor. A pergunta que se pode agora colocar é: será que é possível, a partir do ponto de vista de X, distinguir entre @ e *w*? Tendo em conta que estes dois mundos são idênticos em quase todos os aspetos, X não pode distinguir as duas situações através da perceção do que ocorre ao seu redor, nem a partir das suas disposições para ter estes ou aqueles estados mentais nem para se comportar desta ou daquela maneira. A única hipótese que X tem de distinguir entre @ e *w* passa por ter acesso à informação acerca do estado que ocupa nele próprio o papel causal da dor –

informação relativa à instanciação de propriedades físicas, portanto, dada a teoria da identidade que D. Lewis aceita. Mas esta informação não está disponível para ele sem uma investigação laboriosa, como afirma D. Lewis; e, assim, se assumirmos que X não possui um conhecimento alargado acerca das suas propriedades neuroquímicas, (3) @ e *w* são indistinguíveis do seu ponto de vista. E, em particular, a sua introspeção – se admitirmos este termo como legítimo – é insuficiente para que X saiba se @ ou *w* é o mundo em que habita. Em primeira pessoa, as duas situações são idênticas.

O argumento de D. Lewis contra a rigidez dos termos mentais apresentado na passagem anterior pode ser encarado como um *reductio ad absurdum*. Vamos supor temporariamente que (4) 'dor' é um designador rígido. A partir de (4), temos de admitir que, apesar de (2), (5) em *w*, X não está a sentir dor.[37] Agora, se introduzirmos um princípio intuitivo como:

.[37] Uma alternativa seria a de dizer que 'dor' é um designador rígido que refere uma propriedade extremamente disjuntiva. Por exemplo, 'circular ou triangular' refere rigidamente uma propriedade de todas as coisas que são circulares e triangulares. No caso de 'dor', a expressão designaria em todas as situações possíveis uma propriedade dos estados que desempenham, no seu contexto, o papel causal da dor. Assim, o estado que ocorre tanto no mundo atual como na situação contrafactual seriam dor, não por causa da instabilidade da referência de 'dor', mas pela instabilidade do realizador da propriedade rigidamente designada.

D. Lewis considera que temos a expressão 'estar em dor' (ou 'ter uma dor') para designar essa propriedade (Lewis 1994: 420). No entanto, considera também que esta tem de ser distinguida da propriedade expressa por 'dor'. A razão para isto é que, diz-nos: «[…] this property is not the occupant of the M-role. It cannot occupy that or any other causal role because it is excessively disjunctive, and therefore no events are essentially havings of it. To admit it as causally efficacious would lead to absurd double-counting of causes.» (Lewis 1994: 420) Se admitirmos que a dor é uma propriedade disjuntiva, somos levados a aceitar que se uma ocorrência de dor causa um certo comportamento, então também a ocorrência do estado físico que realiza a dor nesse caso causa igualmente esse comportamento. Estaríamos a falar de dois eventos e não de apenas um, visto que teriam essências muito diferentes: o primeiro pode ter lugar onde quer que algo desempenhe o papel causal de dor, enquanto o segundo apenas tem lugar onde ocorre *aquele* estado físico que naquela situação desempenha esse papel. Assim, o comportamento é causado por dois eventos – e, se quisermos admitir outras disjunções mais artificiais do que a que associamos à dor, teríamos infinitas causas desse comportamento!

Segundo D. Lewis, um evento *c* é causalmente dependente de um outro evento *e* se e só se existe uma dependência contrafactual entre a família de proposições O(*e*), ~O(*e*) e a família O(*c*), ~O(*c*), sendo O(*x*) a proposição que afirma a ocorrência de um evento *x*. Isto é o mesmo que dizer que são verdadeiras as seguintes afirmações contrafactuais:

(1) O(*c*) □→ O(*e*),

(2) ~O(*c*) □→ ~O(*e*).

D. Lewis define a causação em termos de dependência causal, do seguinte modo: se *c*, *d*, *e*, … é uma sequência causal, i. e., se para qualquer *n* o evento que ocupa o lugar *n* + 1 dessa sequência depender causalmente do evento que ocupa o lugar *n* – neste caso, se *d* depende de *c* e *e* depende de *d* –, então para qualquer *n*, o evento que ocupa o lugar *n* dessa sequência causa todos os eventos que ocupem um lugar *m*, tal que *m* > *n* – neste caso, *c* causa *d* e *e*, *d* causa *e*, … (Lewis 1973a: 561-63).

(6) Qualquer sujeito X é capaz de distinguir entre uma situação em que está a ter a experiência consciente E, de uma situação em que não está a ter E,[38]

Deixa de ser verdade que (3), dado que se aceitarmos (3), tendo em conta (5), temos de dizer que existe uma situação contrafactual *w* em que, ao contrário do que acontece em @, X não está a sentir dor, mas mesmo assim X não distingue entre @ e *w*,

---

A partir desta definição, percebemos que a admissão de um evento disjuntivo como a ocorrência da dor «disjuntiva» levaria a admitir a causação múltipla. Por isso, D. Lewis rejeita simplesmente a existência de tais eventos. Ao rejeitá-los é obrigado a abandonar a ideia de que a dor disjuntiva é causalmente eficaz. Existe essa propriedade, mas nenhum evento é essencialmente a ocorrência da mesma, e é por isso que ela deixa de ocupar um lugar em qualquer rede causal. Por isso, a encruzilhada agora é esta: ou seguimos a rigidez de 'dor' em detrimento de assumir que a dor é o ocupante do papel causal atribuído pela psicologia popular, ou admitimos a não rigidez e continuamos a trabalhar com o conceito que já antes tínhamos de dor. A opção de D. Lewis é a de preservar a análise do termo que previamente foi encontrada.

Do mesmo modo que 'dor' é, nesta perspetiva, um designador não rígido, 'a ocorrência de dor em *t*' é uma descrição acidental de um evento. Esta diz respeito a um evento que pode ter lugar sem ser uma ocorrência de dor. É assim que D. Lewis pode admitir que existem eventos como ter uma dor, sem se comprometer com a ideia de que existem eventos disjuntivos, como aqui afirma: «Whenever some term nonrigidly designates the occupant of a role, and that role could be occupied in a variety of ways, the term becomes unsuitable for essential specification of events. If being fragile means having some or another basis for a disposition to break when struck, and if many different properties could serve as such bases (under this- or otherworldly laws), then no genuine event is essentially classifiable as the window's being fragile. There is a genuine event which is accidentally classifiable in terms of fragility, essentially, however, it is a possession of such-and-such molecular structure, that being the actual basis of the window's fragility» (Lewis 1986a: 267).

38 Apesar de intuitivo, o princípio pode acabar por revelar-se falso. Imaginemos o caso em que X está a olhar para uma imagem complexa, recheada de detalhes que requerem uma enorme atenção para serem percebidos. Pode ser verdade que existe uma situação contrafactual em que X está a olhar para uma imagem bastante semelhante, mas com uma pequena parte alterada, e X é incapaz de distinguir a situação atual da contrafactual, e, ainda, que a diferença entre as duas situações envolve diferentes experiências conscientes – ou seja, numa delas X está a ter a experiência E, e na outra não.

Não tenho a certeza de que isto é possível. Podia ser defendido que, se as diferenças não estão a ser capturadas pelo sujeito, não haveria nas duas situações qualquer diferença na experiência consciente de X. Mas não pretendo comprometer-me com nenhuma destas hipóteses; e creio não ser necessário fazê-lo para os propósitos presentes: é suficiente dizer que existe uma classe de experiências conscientes – as experiências fortes, vívidas ou claras, digamos assim – para as quais (6) se aplica, e que algo como uma dor severa faz parte dessa classe. Se assim quisermos, interpretamos o argumento de D. Lewis como estando a aplicar-se concretamente a uma dor severa – o que parece mesmo ser a intenção de D. Lewis, como revelam as suas afirmações anteriormente citadas.

Mas, nesse caso, não estamos também a dizer que o argumento só tem aplicação relativamente a esse tipo de experiência? Isto é, não temos de dizer então que apenas os nomes para experiências fortes *não* são rígidos? Sim, temos. Se pretendermos que o argumento seja a favor de que *nenhum* termo mental pode ser rígido, este falharia assim este propósito – ao reinterpretar (6), teríamos de recuar nas próprias pretensões do argumento. Mas a intenção de D. Lewis pode ser mais modesta: a de mostrar que *nem todos* os termos mentais podem ser rígidos. Mesmo estabelecendo esta conclusão mais fraca, temos em mãos uma razão para preferir à partida analisar os termos mentais como designadores não rígidos.

o que contradiz (6). Assim, por admitirmos (4) e (6), tivemos de rejeitar (3), algo que antes havíamos assumido. Se a balança de plausibilidade pender para (6) em vez de (4) – como D. Lewis parece concordar – conclui-se que 'dor' *não* é um designador rígido.

Outra razão para defender esta conclusão, claro, é o facto de a designação não rígida ser útil para resolver os problemas levantados acima. Estamos a meio do caminho para a solução pretendida. Os problemas colocados anteriormente diziam respeito, recordemos, a quaisquer indivíduos possíveis, atuais ou não. Se o par de indivíduos (ou de conjuntos de indivíduos) problemáticos estiverem localizados em dois mundos possíveis distintos, não há qualquer problema com a identidade entre M = F. Visto que é uma identidade contingente, pode aplicar-se ao mundo de apenas um dos indivíduos – no outro, aplicar-se-á outra identidade, ou nenhuma, se estivermos a lidar com o segundo problema.

Os problemas foram para já estreitados, mas ainda não foram anulados, já que os indivíduos podem habitar o mesmo mundo possível. Mas, segundo D. Lewis, o passo necessário para chegar à solução é pequeno. Consideremos esta passagem:

> «If a nonrigid concept or name applies to different states in different possible cases, it should be no surprise if it also applies to different states in different actual cases. Nonrigidity is to logical space as other relativities are to ordinary space. If the word "pain" designates one state at our actual world and another at a possible world where our counterparts have a different internal structure, then also it may designate one state on Earth and another on Mars. Or, better, since Martians may come here and we may go to Mars, it may designate one state for Earthlings and another for Martians.» (Lewis 1983b: 219)

Esta ideia não é surpreendente, tendo em conta a teoria da modalidade defendida por D. Lewis (*ver* capítulo 1). Aquilo que é verdade em cada mundo é como aquilo que é verdade em cada lugar, momento ou contexto dentro de um único mundo. Admitir a flexibilidade contextual do referente dos conceitos mentais é apenas um pequeno passo, a partir do momento em que é admitido que estes conceitos não são rígidos.

Assim sendo, o valor de verdade da identidade entre M e F pode variar não apenas entre mundos, mas também entre partes de um único mundo. Um mundo meramente possível, recordemos, é como qualquer lugar no mundo atual, mas que está isolado de nós

espáciotemporalmente. Mundos inteiros podem ser qualitativamente idênticos (a nível intrínseco) a uma parte própria de outro mundo, e aquilo que é verdade de um estado mental nesses mundos inteiros é, muito provavelmente, verdade do mesmo estado mental nas partes correspondentes dos outros mundos. Imaginemos dois mundos em que a dor é, em cada um, idêntica a estados físicos diferentes. Há um mundo que contém partes idênticas a esses dois mundos (*ver* secção 1.7). Aquilo que é verdade da dor em cada um desses mundos mais pequenos, é verdade da dor nas correspondentes partes do mundo grande. A contingência dentro de cada mundo é, assim, quase inevitável.

Muito provavelmente, a estratégia mais apropriada não é dividir o mundo atual, e os restantes, em lugares (entendidos como conjuntos de pontos espaciais contíguos) onde vigoram diferentes identidades psicofísicas. Nada nos garante que a realização múltipla não ocorre em indivíduos que habitam a mesma região do mundo. É mais plausível estabelecer uma divisão entre *populações* – i. e., conjuntos (ou somas mereológicas) de indivíduos (alguns deles com apenas um elemento) espalhados por vários lugares do mundo atual, e até por outros mundos possíveis. As populações têm a mesma legitimidade que os lugares a serem tratadas como partes de um mundo (ou do espaço lógico, se admitirmos indivíduos de diferentes mundos).

Não existe uma regra *a priori* para partir o mundo de modo a admitir sem contradição as várias identificações. Assim como só empiricamente podemos descobrir que M = F, também só desse modo podemos vir a saber quão vasta é a população em que essa identidade é verdadeira. Pior do que isso, para saber a extensão dessa população teríamos de conhecer as características fisiológicas de todas as criaturas atuais com mentalidade. Dito isto, aquilo que podemos fazer é formular gradualmente as identidades consoante a evidência disponível.

Não queremos estar obrigados a afirmar que um indivíduo que está em F está também em M, só porque M = F é verdade algures. A relatividade das identidades a populações permite que escapemos a isto. No entanto, é provável que queiramos dizer que um indivíduo numa população tem um certo estado mental, não por ter um estado físico que ocupa o papel causal apropriado, mas porque está integrado numa população em que isso acontece. Este é o caso do louco, um ser humano no qual o disparo de fibras-

C, contrariamente ao que acontece com o resto da espécie humana, não tem as causas e os efeitos que tipicamente associamos à dor (Lewis 1980b: 216). Se aceitarmos que o louco tem dor, então a dor-no-louco não é idêntica ao ocupante de R no louco. Se o louco não for único entre os humanos, teremos também de dizer que dor-em-L, sendo L a população dos loucos, não é o ocupante de R em L.

Para lidar com este caso, D. Lewis introduz a noção de *população adequada* (Lewis 1980b: 219). Se K é a população adequada para um indivíduo X, X está num estado mental M quando X está no estado F que é o ocupante do papel causal adequado a M em K, mesmo que M-em-X não desempenhe esse papel. Suponhamos que X é o louco. Se a sua população adequada relativamente à dor for L, X não está a sentir dor quando nele ocorre o disparo de fibras-C. Pelo contrário, se a sua população adequada for H, a população dos humanos, então o disparo de fibras-C também é a dor em X. Repare-se que X é um elemento de ambas as populações, o que demonstra que o conceito de população adequada é mais estreito que o conceito habitual de população.

Para admitir que o disparo de fibras-C é a dor-em-X temos de supor que, em H, o disparo de fibras-C é a dor-em-H, pelo facto de ocupar o papel causal da dor em H. Mas esta ideia aparece agora sob suspeita pela presença de X em H. Se X pertence a H, então o ocupante do papel causal da dor em H já não é o disparo de fibras-C, visto que isso não se verifica em X. De acordo com D. Lewis, este não é um verdadeiro problema. Como vimos nas secções anteriores, os ocupantes imperfeitos de um papel teórico podem perfeitamente ser merecedores do estatuto de referentes dos termos teóricos. Esta imperfeição pode ser devida (1) ao facto de ocuparem em todas as instâncias um papel ligeiramente diferente do adequado; ou, de uma forma menos drástica, (2) pelo facto de algumas das instâncias colocarem em causa a universalidade das generalizações da teoria, como o caso dos disparos de fibras-C em X. Nas palavras de D. Lewis: «I spoke of the definite syndrome of typical causes and effects […]; that does not mean that it has them invariably. Again, I spoke of a system of states that comes near to realizing commonsense psychology. A state may therefore occupy a role for mankind even if it does not at all occupy the causal role of pain for some mad minority of mankind.» (Lewis 1980b: 219)

Resta agora ser definida a relação de adequação entre uma população e um indivíduo para um certo estado mental. A importância desta relação reside no facto de diferentes associações de indivíduos a populações poderem desencadear diferentes atribuições de estados mentais a um indivíduo. Por exemplo, o louco está a sentir dor se a sua população adequada relativamente à dor for a população dos humanos; mas ele pode estar a sentir prazer se, pelo contrário, for associado a uma população em que o disparo de fibras-C ocupa um diferente papel. É conveniente, por isso, olhar para a própria explicação de D. Lewis:

> «We may say that X is in pain simpliciter if and only if X is in the state that occupies the causal role of pain for the appropriate population. But what is the appropriate population? Perhaps (1) it should be us; after all, it's our concept and our word. On the other hand, if it's X we're talking about, perhaps (2) should be a population that X himself belongs to, and (3) it should preferably be one in which X is not exceptional. Either way, (4) an appropriate population should be a natural kind - a species, perhaps.» (Lewis 1980b: 219)

Temos, então, quatro critérios. Segundo o primeiro, uma população adequada para X relativamente a M deve ser a *nossa* população, aquela que usa o conceito de M e as várias generalizações da psicologia popular para prever e explicar o comportamento dos outros sujeitos. (Mas, afinal, qual população? Visto que é aquela a que pertence o conceito de M, excluímos certas populações a que *nós* pertencemos – a dos seres vivos, a dos animais e a dos primatas. Grande parte dos seus membros não possuem esse conceito. No entanto, permanece ainda indeterminado a que população D. Lewis se refere. Será à espécie humana desde a sua origem até ao final dos tempos? Os humanos do *presente*? Os humanos que existiram até agora? Aqueles que, como D. Lewis, têm o inglês como língua materna?)

Se esta população puder ser a população adequada para X mesmo que X não seja um humano, então a população adequada para X pode não ser um membro do conjunto de populações a que X pertence. No entanto, isto é contradito pelo segundo critério. Este diz-nos que a população deve ser uma a que X pertença. Esta dissonância sugere que, para determinar a população adequada, temos às vezes de dar preferência a um destes

dois critérios. A razão para optar por um deles talvez seja dada pelo terceiro e quarto critérios.

Imaginemos que, em G, a população dos golfinhos, o disparo de fibras-D ocupa o papel causal da dor. Quando em X, um humano, ocorre o disparo de fibras-D, tal não é acompanhado pelas causas e efeitos típicos da dor. No entanto, X pertence à população G + X, que inclui os golfinhos e X, na qual as fibras-D são responsáveis pela dor. Se G + X for a população adequada para X, então X está a sentir dor quando em X ocorre o disparo de fibras-D. Esta possibilidade é bloqueada pelos terceiro e quarto critérios.

Se X não for louco e, por isso, o disparo de fibras-C faz com que se comporte como qualquer outro humano, devemos preferir a população dos humanos, H, à população G + X, para ser eleita como a população adequada para X relativamente à dor. Segundo o terceiro critério, a população adequada deve ser uma em que X não se torna excecional – e sê-lo-ia se o colocássemos em G + X. Por outro lado, de acordo com o quarto critério a população adequada deve ser um tipo natural – ou seja, um corte não arbitrário na classificação dos vários indivíduos. Tal não é o caso de G + X.

Consideramos antes uma objeção de N. Block, de acordo a qual as teorias como a de D. Lewis, que caracterizam os estados mentais em termos de causas e efeitos descritos de acordo com o senso comum, caem inevitavelmente numa perspetiva chauvinista (*ver* secção 2.2). Vimos que D. Lewis consegue responder a esta questão pela maneira como trata as descrições do comportamento que aparecem nas generalizações da psicologia popular. No entanto, existem casos que essa resposta não permite ainda acomodar: principalmente, o dos paralíticos e o dos cérebros em cubas (Block 1980: 283-84).

O primeiro caso pode agora ser resolvido. Tal como acontece com o louco, um paralítico pode estar em M, mesmo que nele M não desempenhe o papel causal adequado: é suficiente que esteja no estado que ocupa esse papel na maioria dos membros da sua população. Quanto aos cérebros numa cuba, a situação é um pouco distinta. Se, para resolver este problema, admitirmos que uma população como a dos humanos é adequada para um cérebro, isso criaria o desconforto de termos de aceitar uma população artificial: nenhuma população formada por humanos e cérebros é plausivelmente um tipo natural, o que cria uma dissonância com o quarto critério considerado acima. A solução que

vislumbro é pôr em ação o primeiro critério, aquele que coloca em relevo os estados mentais dos seres humanos. Isto podia resolver a situação se o cérebro numa cuba for neurológica e quimicamente semelhante ao cérebro humano. No entanto, não vejo como esta solução possa dar-nos uma resposta satisfatória se, em vez de um cérebro, estivermos a falar de um outro órgão ou sistema responsável pela mente dos marcianos que foi isolado do resto do organismo. Por um lado, a simetria entre esta situação e a dos humanos parece indicar que o *cérebro* marciano também tem estados mentais; por outro lado, nenhum critério que consideramos anteriormente parece favorecer esta posição. Talvez tenhamos de concluir que este é um caso problemático para o qual a resposta tem de permanecer indeterminada.

Capítulo 3 – A interpretação radical

3.1 – O projeto de interpretação radical

Em "Radical Interpretation" (1974), D. Lewis começa por apresentar a *interpretação radical* como um projeto (um enigma) prático que consiste em descobrir as atitudes intencionais e a semântica da linguagem de um certo sujeito, unicamente a partir de informação acerca das suas propriedades físicas, como é revelado por estas palavras: «I can diagram the problem of radical interpretation as follows. Given P, the facts about Karl as a physical system, solve for the rest.» (Lewis 1974: 331)[39, 40] Um pouco mais à frente, porém, podemos ver que esta é apenas uma maneira conveniente de expor o problema da interpretação radical, que D. Lewis vê mais como dizendo respeito à questão de como é que as propriedades físicas de um sujeito *determinam* o seu conteúdo mental e semântico. A interpretação radical é, então, uma questão metafísica de como o físico constitui o mental, em vez da questão epistemológico de como o físico é *evidência* para o mental (Lewis 1974: 332-34). Ainda melhor é, creio, dizer que a interpretação radical é um projeto de análise dos conceitos das atitudes intencionais, que consiste em descobrir que condições físicas estabelecem uma aplicação correta desses conceitos.

Ao colocar assim este problema, D. Lewis está deliberadamente a assumir que existe uma relação de superveniência entre as propriedades físicas e mentais de um sujeito. Ele pressupõe que não pode haver dois sujeitos fisicamente idênticos que diferem a nível mental ou semântico (Lewis 1974: 334). Equivalentemente não pode haver qualquer variação a nível intencional sem variação a nível físico, tanto num mundo como num sujeito. A intencionalidade, para D. Lewis, não é um aspeto do mundo independente

39 A expressão 'interpretação radical' remonta ao trabalho de Donald Davidson (*ver* Davidson 1973).

40 Nas discussões relativas a esta classe de estados mentais, D. Lewis realça especialmente as crenças e os desejos como exemplos paradigmáticos. Em (Lewis 1974: 332), D. Lewis mostrou esperança em ser possível reduzir todas as restantes atitudes proposicionais a padrões de crenças e desejos. Essa confiança parece ter-se desvanecido ligeiramente quando nos diz, mais tarde, que «I think it an open question to what extent other states with content – doubting, wondering, fearing, pretending, … – require separate treatment and to what extent they can be reduced to patterns in belief and desire and contentless feeling.» (Lewis 1994: 421) Seja como for, a posição de D. Lewis acerca da interpretação radical é formulada apenas para crenças e desejos.

do padrão de instanciação de propriedades físicas, mas antes um aspeto que está dependente deste último.

É certamente compreensível, assim, esta afirmação de J. R. G. Williams: «David Lewis was no fan of primitive intentionality. He wanted to explain how the intentional – including mental and linguistic representation – could arise in a fundamentally physical world.» (Willliams 2015: 367) Explicar isto é o propósito da interpretação radical. D. Lewis pretende saber o que há em comum a todas as condições físicas que necessariamente estão associadas à presença de uma certa propriedade intencional.

Antes de mais, convém notar que assumir a determinação do mental pelo físico não é, de todo, o mesmo que assumir a impossibilidade de casos de indeterminação quanto à exemplificação de uma propriedade mental por um sujeito. A tese da superveniência implica apenas que a indeterminação mental ou semântica num sujeito tem de estar presente em qualquer outro com exatamente as mesmas propriedades físicas (Lewis 1994: 414). Afirma D. Lewis que «it does not matter that there might be two equally correct ways to resolve some mental or semantic indeterminacy, so long as *both* ways are available for *both* Karls. The two Karls still do not differ.» (Lewis 1974: 334)

Tenho vindo até agora a falar de propriedades físicas. No sentido estrito, estas são as propriedades que dizem respeito às partículas físicas que compõem as coisas e à maneira como as partículas estão estruturadas no espaço e no tempo. Assim concebidas, no entanto, as propriedades físicas são insuficientes para determinar os factos mentais e semânticos de uma pessoa e não podem ser a base de superveniência pretendida. Como vimos antes, D. Lewis considera que aquilo que é definitivo da mentalidade é a ocupação de papeis causais e, como adiante veremos, essa ocupação é crucial na interpretação radical das atitudes e da lingugagem de um sujeito. Ora, duas pessoas possíveis com exatamente as mesmas propriedades físicas podem ter estados físicos a ocuparem diferentes papeis causais – basta para isso que habitem em mundos com diferentes leis da natureza. D. Lewis tem, assim, de ter uma noção mais alargada de propriedades físicas. Aqui apresenta um inventário de aspetos que contam como informação física sobre um sujeito:

> «P, our ultimate data base, gives us the whole truth about Karl as a physical system. It tells us how Karl moves, what forces he exerts on his surroundings, what light or sound or chemical substance he absorbs or emits. It tells us the same things about all of Karl's material parts, great or small, permanent or temporary. It tells us all the masses and charges of the particles that compose him, and all the magnitudes and directions of the fields and potentials and radiation that pervade him. It tells us not only his present physical state but also his physical history; and not only the actual particular physical facts but also the nomic or counterfactual or causal dependences among them. It tells us higher order facts, if need be: as that there exist some or other states of Karl, of unspecified character, that realize such-and-such patterns of causal relations to one another and to such-and-such specified physical states.» (Lewis 1974: 331-32)

A interpretação radical de X tem, então, de ter por base (1) os movimentos de X, (2) a interação de X com o meio circundante próximo (as forças que exerce, as substâncias que emite e que absorve, por exemplo), (3) as propriedades físicas de todos os constituintes de X, (4) a história passada de X relativa a factos do mesmo tipo que (1), (2) e (3), (5) as disposições de X e as dependências causais entre os constituintes de X, descritas através de afirmações contrafactuais e, finalmente, (6) as interações causais entre aquilo que foi descrito em (1), (2), (3) e (4).[41]

A interpretação radical, para D. Lewis, é indissociável da análise dos nomes para estados mentais através de papeis causais. É analiticamente verdade, vamos supor, que uma experiência de vermelho é um estado normalmente causado por objetos vermelhos. Quando é que um sujeito tem uma experiência de vermelho? Obviamente, quando tem um estado que normalmente é causado por objetos vermelhos. Agora, queremos saber quando é que um sujeito tem esta ou aquela atitude. Mais uma vez, a resposta é: quando o sujeito tem um estado que ocupa o papel causal característico dessa atitude.

---

41 Ao incluirmos propriedades contrafactuais na classe de propriedades físicas, deixamos de poder dizer que as propriedades físicas são intrínsecas. A superveniência não ocorre, por isso, unicamente entre algumas propriedades intrínsecas básicas e as propriedades mentais – há que ter em conta aspetos que estão fora dos limites do corpo do sujeito. Isto é, pelo menos, o que acontece de acordo com D. Lewis, tendo em conta que ele defende a hipótese de que as propriedades contrafactuais e causais, assim como as leis da natureza, dependem do padrão total de instanciação de propriedades físicas – ou propriedades semelhantes às físicas – num mundo (*ver* Lewis 1986c: vii-xv). Por isso, as propriedades causais locais de um sujeito dependem do que acontece no resto do mundo. Mas podemos, pelo menos, dizer algo deste género: as propriedades mentais são determinadas pelas propriedades físicas intrínsecas de um sujeito e pelas relações contrafactuais que se verificam no mundo. Esta conjunção de propriedades é simplesmente a conceção lata de propriedades físicas com a qual estamos a trabalhar desde o início deste capítulo.

Contudo, a situação agora é mais complexa, devido à natureza dos estados mentais que estamos a considerar. Queremos atribuir a X certas crenças, desejos, intenções, medos, entre outras atitudes, mas a psicologia popular não nos fornece diretamente uma descrição funcional, por exemplo, da crença de que está a chover ou do desejo de que o gato não fuja. Como foi visto no capítulo anterior (*ver* secção 2.2), a psicologia popular estabelece o papel causal das crenças, dos desejos, das intenções, dos medos, entre outras, em geral, abstraídos de qualquer conteúdo. Felizmente, podemos facilmente chegar a princípios relativos às atitudes indexadas com um conteúdo a partir dos princípios gerais. Vimos antes que estes princípios envolvem a quantificação universal sobre proposições e descrições de estímulos e comportamento que ocorrem no âmbito da quantificação. A partir destes, podemos simplesmente formar a frase aberta que resulta de eliminar o quantificador que liga as variáveis proposicionais, e substituir as variáveis por nomes para proposições.

Assim, temos agora os meios para esboçar a solução ao problema da interpretação radical. Tendo o papel causal de cada um dos pares ordenados (A, P) de atitudes e proposições, podemos simplesmente ver qual é o conjunto desses pares que melhor se enquadra nos papeis causais dos estados físicos de X. É exatamente isso o que aqui nos diz D. Lewis: «What are the constraints by which the problem of radical interpretation is to be solved? Roughly speaking, they are the fundamental principles of our general theory of persons. They tell us how beliefs and desires and meanings are normally related to one another, to behavioral output, and to sensory input.» (Lewis 1974: 334) (A mesma ideia pode ser encontrada, por exemplo, em (Lewis 1979a: 533, 1983a: 5, nota 2, 1983b: 373 e 1994: 421).)

É necessário introduzir agora uma modificação neste esquema. D. Lewis considera que o louco está a sentir dor, mesmo não tendo qualquer estado físico que desempenha o papel característico da dor (*ver* secção 2.5). Da mesma maneira, um qualquer outro sujeito pode ter a crença ou o desejo de que P, apesar de nenhum dos seus estados se conformar ao papel que é para ele especificado indiretamente pela psicologia popular. Relembro que a explicação para isto é o o facto de o louco estar num estado físico F que, na população K, ocupava na maior parte dos casos o papel causal da dor, e,

ainda, de K ser a população adequada para o louco (é um tipo natural a que ele pertence, não há uma população alternativa, mas igualmente natural, relativamente à qual ele deixaria de ser louco, e por aí em diante). Esta possibilidade leva a algumas complicações no método de interpretação, como aqui explica D. Lewis:

> «An interpretation just of Karl at the present moment need only specify his attitudes and his meanings. But an interpretation of Karl's kind generally – or even of Karl himself as he is at various times, or as he might have been under various different circumstances – must be something more complicated. It must be a scheme of interpretation specifying the attitudes and meanings as a function of the momentary total physical state. On the basis of such states, the scheme assigns interpretation to individuals at times. (Indeed it might – and should, I think – do this simply by identifying certain attitudes with certain (partial) physical states.) The best scheme is the one that does the best job overall of conforming to the constraining principles, taking one individual and time with another. (The individuals in question being not only Karl and others of his kind as they actually are, but also some of their might-have-been counterparts.)» (Lewis 1983c: 119-20)

O método de interpretação radical de um indivíduo X passa agora a ser desdobrado em dois momentos: (1) em cada população K, associa-se cada F (o momentâneo estado físico de um indivíduo) a um conjunto de pares ordenados de atitudes e proposições (A, P). Essa função deve ser estabelecida tendo em conta a adequação de cada F nos vários indivíduos de cada K aos papeis causais especificados pela psicologia popular para as várias atitudes, de um modo análogo àquele que anteriormente foi apresentado para interpretar X individualmente.[42] E, finalmente, (2) a interpretação correta de X no momento *t* é aquela que está associada, na população adequada a X, ao estado físico F em que X se encontra em *t*. Sendo D. Lewis um teórico da identidade, considera que a função que liga cada F a certas atitudes pode ser compreendida como especificando a identidade

[42] Visto que, agora, estamos a tratar de um conjunto de indivíduos, em vez de cada um deles isoladamente, podemos dispensar falar das propriedades contrafactuais de cada indivíduo X e dos seus vários constituintes e, como D. Lewis faz nesta passagem, incluir nesse conjunto de indivíduos que estamos a considerar, alguns que pertencem a outros mundos possíveis próximos. Esta transição é justificada pelo facto de D. Lewis considerar que a verdade de uma afirmação contrafactual deve ser avaliada pela semelhança qualitativa entre mundos possíveis (Lewis 1973b). Ora, se aquilo que se está à procura é de um esquema de interpretação que seja adequado não só relativamente àquilo que ocorre atualmente aos indivíduos de uma população, mas também àquilo que ocorreria em certas circunstâncias contrafactuais, passa a ser o mesmo especificar as disposições dos vários indivíduos atuais ou incluir no grupo de indivíduos considerado aqueles que habitam os mundos próximos relevantes.

– talvez não de F, mas de estados parciais de F – com uma atitude ou com um certo sistema de atitudes.

Com isto, deixamos de poder dizer que as propriedades físicas, no sentido lato, determinam as propriedades mentais de X, visto que estas últimas podem depender daquilo que ocorre nos outros membros da população de X. Talvez possamos enfraquecer a hipótese mantendo-a ainda assim interessante, dizendo que a relação de superveniência é estabelecida entre as propriedades físicas de todos os indivíduos da população a que X pertence. Se, localizado num qualquer mundo possível distante, Y fizer parte de uma população com exatamente as mesmas propriedades que a população de X, podemos estar seguros de que X e Y são duplicados mentais.

3.2 – As atitudes e a psicologia popular

Além de esboçar uma solução para o problema da interpretação radical, D. Lewis sugere mesmo alguns princípios reguladores da atribuição de conteúdo mental que considera estarem presentes na psicologia popular. Em "Reduction of Mind" (Lewis 1994), contrasta aquilo que tem a dizer acerca disto com as teses de um imaginário filósofo chamado Espantalho:

> «Strawman says that folk-psychology says – and truly – that there is a language of thought. It has words, and it has syntactic constructions whereby those words can be combined into sentences. Some of these sentences have a special status. Strawman says they are 'written in the belief box' or 'in the desire box', but even Strawman doesn´t take that altogether literally. There are folk-psychological causal roles for words, for the syntactic constructions, and for the belief and desire boxes. It is by occupying these roles that the occupants deserve their folk-psychological names.
> The question what determines content then becomes the question: what determines the semantics of the language of thought? Strawman says that folk psychology specifies the semantic operations that correspond to syntactic constructions such as predication. As for the words, Strawman says that folk psychology includes, in its usual tacit and unsystematic way, a causal theory of reference […]. There are many relations of acquaintance that connect the mind to things, including properties and relations, in the external world. […]
> Once the words of the language of thought have their referents, the sentences have their meanings. These are structures built up from the referents of the words in a

way that mirrors the syntactic constructions of the sentences from the words.» (Lewis 1994: 422-23)

Ora, de acordo com a abordagem do Espantalho, X tem a crença (ou o desejo) de que P se e só se (1) X tem uma linguagem de pensamento com regras sintáticas e semânticas que são especificadas pelos princípios da psicologia popular, (2) algumas expressões – paradigmaticamente (e talvez unicamente) os predicados e os nomes – da linguagem de pensamento de X adquirem valores semânticos através de relações com itens do ambiente externo, (3) outras expressões, como quantificadores, variáveis e conectivos lógicos, têm um valor semântico especificado pela psicologia popular, (4) P é uma proposição estruturada formada a partir de valores semânticos das expressões da linguagem de pensamento de X. Se assumirmos que P é uma qualquer proposição singular (F, A), então (5) existe um predicado da linguagem de pensamento de X, 'F', que tem como valor semântico a propriedade F, (6) existe um nome dessa linguagem, 'A', que tem como valor semântico o objeto A, (7) as regras sintáticas da linguagem de X permitem construir uma frase, 'F(A)', que tem como significado a proposição (F, A), e, por fim, (8) a frase 'F(A)' desempenha no *hardware* de X um papel causal a que podemos chamar "estar na caixa das crenças (ou desejos)". ((5)-(8) podem ser adaptadas para frases envolvendo quantificadores, variáveis e conectivos lógicos.)

Esta abordagem trata o problema da interpretação das atitudes de X como secundário relativamente ao problema de especificar a semântica de LP (a linguagem de pensamento de X). Assim, os princípios da psicologia popular que têm algo a dizer sobre as atitudes são essencialmente aqueles que dizem respeito à semântica de uma linguagem como LP e aos papeis causais das frases de LP em X. A abordagem de D. Lewis é diametralmente oposta a esta, visto que começa por fornecer uma teoria das atitudes de X, e apenas a partir destas considerar as propriedades semânticas da linguagem de X.

Antes de passarmos a considerar a proposta positiva de D. Lewis, é importante ter em conta as objeções que encontra para refutar uma abordagem como a do Espantalho. Uma delas é a seguinte. D. Lewis considera que a psicologia popular nada tem a dizer acerca do modo como as crenças e os desejos são implementados fisicamente (Lewis 1994: 422). Isto não é o mesmo que dizer que a hipótese da linguagem de pensamento é

falsa. Pelo contrário, é dizer que, contrariamente aos princípios da psicologia popular considerados em conjunto, essa hipótese tem um estatuto meramente empírico e contingente. É por isso que não deve entrar na definição das atitudes intencionais. Como afirma D. Lewis num artigo anterior: «It may be true that men think in language, and that to hold a belief is to accept a sentence of one's language. But it does not follow that belief should be analyzed as acceptance of sentences. It should not be. Even if men do in fact think in language, they might not. It is at least possible that men – like beasts – might hold beliefs otherwise than by accepting sentences.» (Lewis 1975: 27)[43]

Uma das hipóteses que D. Lewis deixa em aberto é a de que talvez não se encontre um estado físico em X que corresponda exatamente a cada uma das suas crenças e a cada um dos seus desejos. Para explicar esta possibilidade, recorre à analogia entre um sistema de crenças e um mapa. A transmissão de informação através de um mapa não envolve a apresentação de certos traços – como as frases de uma linguagem – que expressam um determinado conteúdo independentemente dos outros traços do mapa. Como D. Lewis explica:

> «Suppose I have a piece of paper according to which, *inter alia*, Collingwood is east of Fitzroy. Can I tear the paper up so that I get one snippet that has exactly the content that Collingwood is east of Fitzroy, nothing more and nothing less? If the paper is covered with writing, maybe I can; for maybe 'Collingwood is east of Fitzroy' is one of the sentences written there. But if the paper is a map, any snippet according to which Collingwood is east of Fitzroy will be a snippet according to which more is true besides. For instance, I see no way to lose the information that they are adjacent, and that a street runs along the border. And I see no way to lose all information about their size and shape.» (Lewis 1994: 422)

43 Nesta passagem, D. Lewis pode não estar a referir-se especificamente a uma linguagem de pensamento, mas a uma linguagem pública. Estas considerações são a resposta a uma possível objeção à abordagem do significado da linguagem de uma população fazendo referência às atitudes dos seus membros. A objeção é assim apresentada: «It is circular to define the meaning in P of sentences in terms of the beliefs held by members of P. For presumably the members of P think in their language. For instance, they hold beliefs by accepting suitable sentences of their language. If we do not already know the meaning in P of a sentence, we do not know what belief a member of P would hold by accepting that sentence.» (Lewis 1975: 27) Mas, seja como for, a resposta de D. Lewis aplica-se igualmente à hipótese da linguagem de pensamento. Além do mais, mostra-nos também que a abordagem de D. Lewis à atribuição de atitudes é contrária a qualquer posição que tente analisar as crenças e os desejos como a aceitação e negação de frases de uma linguagem pública (ou talvez, a disposição para as afirmar ou negar).

Neste sentido, não podemos dizer que existe uma parte do mapa descritível como aquela que expressa única e exclusivamente que P. Existe, obviamente, uma parte do mapa que transmite a informação de que P, mas também de que Q, R, ..., sendo estas proposições diferentes de P. Pelo contrário, as frases permitem expressar que P isoladamente. Num mapa, a informação presente nas várias partes interliga-se; num texto, esse fenómeno é pelo menos evitável. Se, como defende o Espantalho, a crença de que P é idêntica a uma expressão linguística, existe um estado físico que expressa isoladamente P; mas isso não acontece se as crenças forem mais próximas do mapa e puderem apenas ser consideradas como um sistema de várias crenças: de que P, Q, R, ….

Outras duas objeções dizem respeito àquilo que uma abordagem deste género não consegue acomodar. Por um lado, D. Lewis considera haver uma distinção importante a fazer entre o conteúdo *restrito* e o conteúdo *lato* (ou *de re*). O primeiro deve ser determinado unicamente por aquilo que se passa com o sujeito, enquanto que o último envolve ainda as coisas com que o sujeito se confronta, sobre as quais tem esta ou aquela conceção, entre outras coisas. Duas pessoas podem ter atitudes com exatamente o mesmo conteúdo restrito e diferente conteúdo lato, como demonstra o famoso caso da Terra Gémea; e vice-versa, como é exemplificado pelo caso de Pierre (*ver* capítulo 4). Se o Espantalho estiver correto, no entanto, estas diferenças desvanecem-se, pelo facto de o conteúdo de uma grande e importante parte das expressões da linguagem de pensamento terem como conteúdo exatamente os itens que a elas estão ligadas por uma certa cadeia causal, ou seja, terem apenas conteúdo lato (Lewis 1994: 423-25). Por outro lado, o Espantalho não pode, de acordo com D. Lewis, acomodar também o conteúdo irredutivelmente *egocêntrico*, aquele que diz respeito ao próprio sujeito e não ao mundo (*ver* capítulo 4). Por exemplo, como é que poderia existir uma distinção entre a crença de X de que ele próprio vive em Itália e a de que X vive em Itália? Talvez a sua linguagem de pensamento contenha uma palavra análoga ao ‘eu’ da linguagem pública, mas, nesse caso, o valor semântico de «eu» devia ser X, assim como aconteceria como uma qualquer expressão sinónima de ‘X’ na sua linguagem de pensamento. Por isso, em ambos os casos

a proposição em que X acredita é exatamente a mesma, algo como (vive em Itália, X) (Lewis 1994: 425-27).[44]

A última objeção é que a abordagem do Espantalho não deixa qualquer lugar à racionalidade como critério de atribuição de conteúdo. De acordo com D. Lewis, alguns dos princípios da psicologia popular afirmam que os sujeitos das atitudes são, em grande medida, racionais e inteligíveis. Um desses princípios de racionalidade diz-nos que optamos pelo comportamento que serve da melhor maneira os nossos desejos, de acordo com as nossas crenças (Lewis 1974: 337-38, 1986b: 36-7, 1994: 427). Outro desses princípios diz-nos que o nosso sistema de crenças – e, consequentemente, o nosso sistema de desejos instrumentais – muda constantemente de uma maneira coerente com a evidência de que dispomos através da experiência, e que grande parte das nossas crenças mais próximas da experiência percetiva são em grande medida verdadeiras (Lewis 1994: 427).[45]

Antes de considerarmos a contribuição destes princípios para a interpretação radical, é conveniente ter em conta a formulação de cada um deles em termos mais rigorosos, de uma maneira proposta por D. Lewis em vários dos seus artigos. A primeira

---

44 Não estou certo de que o Espantalho não pudesse realmente fazer melhor neste particular. Creio que a questão depende em parte daquilo que se considera como a linguagem de pensamento de X. Se for uma linguagem que X partilha com os elementos da sua espécie, por exemplo, podíamos dizer que existe uma palavra como 'eu' que desempenha em todos eles o papel causal de entrar em crenças que são irredutivelmente acerca dos próprios sujeitos. De facto, se a linguagem for apenas a do próprio X, a ocorrência da palavra 'eu' e de 'X' parece não apresentar quaisquer diferenças, mas ao ser alargada a outros pensadores, a situação podia ser diferente. O valor semântico de 'eu' seria talvez mais corretamente descrito como uma função *f* extremamente simples: para cada sujeito X, *f*(X) = X. Talvez assim se pudesse caracterizar o conteúdo egocêntrico de X como sendo algo como (F, *f*(*x*)), em vez de (F, X), como parece ser aquilo que D. Lewis considera que é a única solução para o Espantalho.

45 Este princípio pode ser entendido como um princípio *mínimo* de caridade. Não se aproxima, no entanto, do princípio de caridade apresentado por D. Lewis em (Lewis 1974: 336-37). Aquele com que estamos agora a lidar envolve a caridade interpretativa apenas ao nível da verdade das crenças percetivas de um sujeito e relativamente ao facto de se admitir que existe uma modificação coerente do resto do sistema de crenças de acordo com a nova informação que lhe chega. Pelo contrário, o princípio de caridade encontrado em (Lewis 1974) vai mais longe, e tem algo a dizer acerca do modo como o sujeito muda as suas crenças de acordo com a experiência percetiva – ou seja, relativamente àquilo que ele acha plausível dada certa informação parcial sobre o mundo. Segundo este princípio mais forte, o sujeito terá mais ou menos as crenças que *nós* – os intérpretes – consideramos adequadas para a evidência que ele teve disponível durante a sua vida. Esta versão do princípio da caridade, contrariamente a versões menos sofisticadas, admite o erro de um sujeito no caso de ele ter sido confrontado com evidência enganadora ou, por exemplo, ter recebido aprendizagem imperfeita (Lewis 1974: 336). Não me parece que D. Lewis alguma vez tenha abandonado este princípio da caridade apresentado em (Lewis 1974). Mas, ao longo do seu percurso, foi tornando-o mais forte, acrescentando novos elementos que serão considerados na próxima secção. Por isso, não me vou debruçar atentamente sobre este princípio para já.

generalização pode ser expressa através do vocabulário da teoria da decisão racional, como é proposto em (Lewis 1974: 337-38), e, por sua vez, a segunda pode ser apresentada em termos da teoria da probabilidade, de uma maneira esboçada em (Lewis 1983b: 374).

Para tratarmos as crenças de um sujeito X introduzimos uma função C, que distribui por cada mundo possível uma parcela do valor máximo de probabilidade – normalmente, esse valor é 1.[46] O valor numérico atribuído por C a cada mundo expressa a probabilidade desse mundo ser o atual de acordo com X. Ao fazer isto, C determina uma probabilidade para cada proposição P, calculada de acordo com a seguinte fórmula:

$$C(P) =_{df} \sum_{w \in P} C(w)$$[47, 48]

(Lewis 1981a: 6)

Se C(P) = 0.7, por exemplo, e se C for adequada para descrever as crenças de X num instante de tempo *t,* dizemos que X acredita com grau 0.7, em *t,* que P. Esta estratégia permite expressar informação relativa à parcialidade das crenças, algo que D. Lewis considera necessário para descrever corretamente as propriedades mentais de um sujeito (Lewis 1974: 333-34, 1986b: 30). Agora, em vez de falarmos em crenças indexadas por um conteúdo proposicional, podemos falar de crenças indexadas por graus e conteúdos. Em vez de pares, têm de ser tratadas como triplos (C, *g*. P). (Acontece o mesmo com os desejos.)

Um problema que se pode agora colocar é que os valores usados numa distribuição de probabilidades deste género são demasiado abundantes para captarem as diferenças entre a força das crenças de indivíduos como os seres humanos, tendo em conta que o detalhe que permitem é muito maior do que aquele que existe no que pretendem descrever. D. Lewis reconhece essa inadequação, mas resolva-a assim:

---

46 A utilização da letra 'C' para nomear esta função tem origem no termo inglês "credence", assim como é utilizado em (Lewis 1981a).

47 Estou aqui a assumir que uma proposição é um conjunto de mundos possíveis, como vimos na secção 1.5.5. No entanto, esta fórmula e tudo o que vai ser dito adiante aplica-se mesmo se quisermos pensar nas proposições de uma outra maneira, desde que esteja garantido que há identidade entre proposições verdadeiras exatamente nos mesmos mundos.

48 Fica aqui o exemplo de um cálculo feito com base nesta fórmula. Vamos supor que P é verdadeira apenas nos mundos w_1, w_2 e w_3, e que $C(w_1) = 0.2$, $C(w_2) = 0.7$ e $C(w_3) = 0.02$. C(P) é a soma dos valores de probabilidade atribuídos a w_1, w_2 e w_3. Ora, com estes valores, C(P) = 0.92.

> «Precise numerical degrees of belief look artificial, so we might favour coarser-grained systems with small number of distinct grades of belief. But whatever small number of grades we took, it is likely that our scale would seem sometimes too coarse to capture real distinctions and sometimes too fine to be realistic. A better response is to continue to treat a belief system as a precise numerical probability distribution, but then to say that normally there is no fully determinate fact of the matter about exactly which belief system someone has.» (Lewis 1986b: 30)

Ou seja, D. Lewis considera que a inadequação verificada numa escala com uma grande quantidade de valores pode por vezes ser verificada também numa outra escala que use um número muito mais reduzido de valores, acrescentando a isso o facto de estas últimas poderem revelar-se simplistas de mais para lidarem com certos casos. D. Lewis propõe, então, que continuemos a trabalhar com distribuições de probabilidade que atribuem valores precisos e, assim que a inadequação começar, tratemos o caso como um exemplo de indeterminação. Assim, em vez de uma, várias funções – C, C1, C2, … – podem ser adequadas para descrever as crenças de X.

D. Lewis propõe que tratemos C como capturando as disposições que um sujeito X tem, num determinado momento, para passar a ter certas crenças de acordo com a evidência disponível. Se, nesse momento, a sua experiência é evidência de que a proposição E é verdadeira, podemos dizer que o sistema de crenças de X passa a ser dado pela função C(-/E). Assim, a probabilidade que X atribui, quando confrontado com a evidência de que E, à possibilidade do mundo *w* ser o mundo atual é $C(w/E)$.[49, 50] Podemos

---

[49] $C(w/E)$ consiste na probabilidade condicional de *w* ser o mundo atual, dado E, de acordo com C. É definida como $C(w \cap E)/C(E)$. Suponhamos que E é o conjunto cujos membros são w_1, w_2 e w_3; e que $C(w_1) = 0{,}2$, $C(w_2) = 0{,}05$ e $C(w_3) = 0{,}15$. Deste modo, $C(E) = 0{,}4$ e, por exemplo, $C(w_1/E) = C(w_1 \cap E)/0{,}4 = C(w_1)/0{,}4 = 0{,}2/0{,}4 = 0{,}5$. Visto que w_1 é um membro de E, a probabilidade condicional de w_1 dado E é mais elevada do que a probabilidade *a priori* de w_1.

[50] Com isto, a função C(-/E), tal como a função C, também atribui valores de probabilidade às várias proposições. Desta vez, isso acontece desta maneira:

$C(P/E) = \sum_{w \in P} C(w/E)$

Ou, equivalentemente:

$C(P/E) = \sum_{w \in (P \cap E)} C(w)/C(E)$

Por exemplo, seja P a proposição cujos membros são w_1, w_2, w_3, w_4 e w_5, sendo que C atribui a todos estes mundos a probabilidade 0,02; seja E a proposição cujos membros são w_3, w_4, w_5, w_6 e w_7, sendo que C aos últimos dois mundos a probabilidade 0,05. Ora, $C(P) = 0{,}1$ e $C(E) = 0{,}16$. Por sua vez, $C(P/E) = C(P \cap E)/C(E) = (C(w_3) + C(w_4) + C(w_5))/C(E) = 0{,}6/0{,}16 = 0{,}375$.

A proposição E vai receber a probabilidade 1 de acordo com a função C(-/E). Ora, repare-se que $C(E/E) = C(E \cap E)/C(E) = C(E)/C(E) = 1$ (seja qual for a probabilidade de E!). Se nenhum dos mundos que

pensar na função C como expressando as crenças que X tinha antes de se confrontar com uma certa experiência percetiva – ou seja, C é a distribuição de probabilidades atribuídas por X em *t*, e C(-/E) é aquela que é atribuída em *t* + 1, sendo que, em *t* + 1, X é confrontado com a evidência de que E. O princípio de que um sujeito muda as suas crenças conforme a evidência que tem disponível consiste exatamente em dizer que C(-/E) é uma função adequada para descrever as probabilidades que atribui aos vários mundos e, consequentemente, às várias proposições, se E for a evidência de que dispõe num determinado instante e C expressar as crenças que tinha imediatamente antes.

Quanto ao primeiro princípio, D. Lewis formula-o assim em termos da teoria da decisão:

> «Take a suitable set of mutually exclusive and jointly exhaustive propositions about Karl's behavior at any given time; of these alternatives, the one that comes true according to P [a informação relativa às propriedades físicas de quem queremos interpretar] should be the one (or: one of the ones) with maximum expected utility according to the total system of beliefs and desires ascribed to Karl at that time […].» (Lewis 1974: 337)

A avaliação da utilidade esperada de um certo comportamento para um sujeito tem de ter em conta tanto as suas crenças como os seus desejos. Tendo já sido capturadas as crenças de X em *t* através da função C(-/E), resta-nos apresentar a informação relativa aos desejos de X em *t* de uma maneira que seja tratável pelos recursos da teoria de decisão. Os desejos de X podem ser dados por uma função V que atribui a cada mundo um número positivo ou negativo que expressa, respetivamente, a desejabilidade ou indesejabilidade de um mundo assim para X. O valor de desejabilidade de uma proposição P para X em *t* é calculdado através da seguinte fórmula:

$$V(P) = \sum_w C(w/P)\ V(w)$$

(Lewis 1981a: 6)

é um elemento de E for membro de uma proposição Q, Q vai receber a probabilidade 0, de acordo com C(-/E), já que $C(Q \cap E) = 0$.

Podemos, assim, calcular o valor atribuído a cada proposição que diga respeito a uma alternativa de comportamento disponível para X em *t*. Este valor, no entanto, de acordo com D. Lewis, não é o mesmo que a utilidade esperada desse comportamento. Calcular esta última envolve ter em conta proposições relativas ao modo como as várias opções influenciam ou não causalmente o que acontece no mundo. A fórmula proposta por D. Lewis é a seguinte, para cada $B \in A$ e cada *k* que pertence a um conjunto de hipóteses de dependência K, sendo B uma proposição relativa a uma alternativa de comportamento disponível para X e A o conjunto exaustivo de proposições desse género mutuamente exclusivas:

$$U(B) = \sum_{k \in K} C(k)\, V(B \cap k)$$[51]

[51] D. Lewis considera adequado fazer uma distinção entre U(B) e V(B) de modo a resolver casos como os problemas de Newcomb, apresentados em (Nozick 1970). Imaginemos este cenário: «*P* knows that *S* or *T* is his father, but he does not know which one is. *S* died of some terrible inherited disease, and *T* did not. It is known that this disease is genetically dominant, and that *P*'s mother did not have it, and that *S* did not have the recessive gene. If *S* is his father, *P* will die of this disease; if *T* is his father, *P* will not die of this disease. Furthermore, there is a well-confirmed theory available, let us imagine, about the genetic transmission of the tendency to decide to do acts which form part of an intellectual life. This tendency is genetically dominant. *S* had this tendency (and did not have the recessive gene), *T* did not, and *P*'s mother did not. *P* is now deciding whether (a) to go to graduate school and then teach, or (b) to become a professional baseball player. He prefers (though not enormously the life of an academic to that of a professional athlete» (Nozick 1970: 125).

P tem duas alternativas a considerar, expressas pelas proposições A1 e A2. A1 é a proposição que nos diz que *P* optou por uma vida académica, e A2 é a proposição segundo a qual *P* optou por uma carreira desportiva. Apesar de entre quaisquer mundos idênticos em tudo exceto quanto ao facto de ser verdade A1 ou A2, *P* preferir aquele em que A1 é o caso, considerando o valor de cada uma delas é plausível que V(A1) < V(A2). Repare-se que V(A1) é calculado tendo em conta a soma das probabilidades de cada mundo ser o atual dado A1 multiplicada pelo valor atribuído a cada mundo; e o mesmo acontece para V(A2). Ora, dado que A1, a soma das probabilidades do mundo em que *P* tem a doença terrível é, de acordo com as suas crenças, maior do que a daqueles em que ele não as tem, e o valor atribuído aos vários mundos da primeira classe é em geral mais pequeno do que aquele que é atribuído aos mundos da segunda classe, já que *P* não deseja de maneira alguma ter a doença. Já A2 vai favorecer as probabilidades dos mundos que são mais desejáveis para *P*, aqueles em que ele não tem a doença.

Se identificarmos o valor dado pela função V com a utilidade esperada, dizemos que o curso de ação mais racional para *P* é escolher A2. Ora, D. Lewis considera que isto é inaceitável. O valor que *P* atribui a A1 e A2 depende do facto de estas alternativas serem ou não o indício de algo – neste caso, uma doença terrível – para a qual nenhuma delas contribui (Lewis 1981a: 8-14). Se considerarmos a utilidade esperada como tendo de considerar as hipóteses de dependência daquilo que ocorre no mundo relativamente às nossas opções, o problema desvanece-se. Segundo a fórmula proposta por D. Lewis, há a considerar a probabilidade de cada uma das hipóteses de dependência multiplicada pelo valor da proposição que consiste na conjunção da proposição que expressa a alternativa e da hipótese de dependência. Ora, a hipótese de dependência a que *P* atribui maior probabilidade é K1, aquela segundo a qual é indiferente se ele torna verdadeira a proposição A1 ou A2 para o facto de ele ter ou não a doença terrível. Comparando V(K1 & A1) e V(K1 & A2), verifica-se que o primeiro é mais elevado. Para cada mundo que é um membro da interseção entre K1 e A1 em que a doença horrível acontece a *P*, existe um mundo idêntico excetuando no

(Lewis 1981a: 11-2)

É assim que as funções C e V atribuídas a X determinam, dada a evidência E, a disposição de X para modificar as suas crenças quando confrontado com E, e o comportamento que, de entre as alternativas ao dispor de X, é mais racional tendo em conta as suas crenças e desejos dado E. Transpondo isto para o âmbito da interpretação radical, devemos dizer, admitindo a verdade dos dois princípios que temos estado a considerar, que um estado físico F deve ser interpretado para a população K como o sistema de crenças e desejos expresso pelas funções C(-/E) e V se e só se, na maioria dos membros de K, (1) F é tipicamente causado por estímulos percetivos que transmitem a informação de que a proposição E é verdadeira em pessoas que estejam num outro estado físico que, por sua vez, é interpretado como sendo o sistema de crenças e desejos expresso pelas funções C e V,[52] (2) F causa tipicamente o comportamento ao qual é atribuído a maior utilidade esperada por C(-/E) e V, e (3) quando um sujeito está em F e é confrontado com estímulos que carregam a evidência de que G, passa a estar num estado físico que deve ser interpretado parcialmente como o sistema de crenças C((-/E)/G). Se quisermos identificar certos estados físicos com crenças e desejos específicos, a situação torna-se mais complexa. Para considerarmos, por exemplo, um estado parcial de F, F1, como a atitude (A, *g*, P), é necessário comparar a contribuição causal de F1 para os estados físicos mais complexos de que este faz parte, com a contribuição de (A, *g*, P) para os vários sistemas de crenças e desejos. O papel causal de F1 e o papel lógico de (A, *g*, P) têm de se espelhar mutuamente.

facto de *P* tornar verdadeiro A2 em vez de A1, que é membro da interseção entre K1 e A2, em que a doença também ocorre. De entre cada par de mundos idênticos em tudo exceto no facto de ser verdade A1 ou A2, *P* vai preferir A1, e por isso é que V(K1 & A1) é mais alto que V(K1 & A2).

52 Ao explicitar as cláusulas (1), (2) e (3), teve de ser feita referência a estados físicos que já receberam interpretação. A interpretação destes, possivelmente, fez também referência à interpretação de F. Isto mostra que um método de interpretação radical como o de D. Lewis é holista: tem de trabalhar imediatamente, desde início, com vários estados e as relações causais entre eles.

3.3 – Hilary Putnam e a indeterminação radical

Em "New Work for a Theory of Universals" (1983b), D. Lewis reconhece que os princípios que acabámos de considerar na secção anterior não são suficientes para levar a cabo a interpretação radical. O problema que ele encontra é uma adaptação do problema que Hilary Putnam apresenta em *Reason, Truth and History* (1981) relativamente a qualquer atribuição de valores semânticos (quer extensões quer intensões) às expressões de uma linguagem, unicamente a partir de restrições que determinam o valor de verdade de frases completas (Putnam 1981: 32-5). Mais concretamente, o problema, defende D. Lewis acerca da mente e H. Putnam acerca da linguagem, é que mais do que uma interpretação é adequada às restrições impostas. Assim sendo, os factos relativos ao conteúdo mental e linguístico ficam radicalmente indeterminados.

Consideremos primeiro aquilo que H. Putnam tem a dizer. Suponhamos que temos em mãos uma série de restrições que determinam o valor de verdade das frases de uma linguagem L, tanto no mundo atual como em qualquer outro mundo possível. Deste modo, fica associada a cada frase de L uma intensão (i. e., uma função de mundos para valores de verdade). É esta intensão suficiente para determinar a interpretação adequada dos componentes das frases de L? Consideremos a frase apresentada como exemplo em (Putnam 1981: 34-5):

(1) Um gato está num tapete.

A intensão desta frase é uma função que atribui o valor 1 (que representa a verdade) a todos os mundos em que um gato está num tapete, e 0 (que representa a falsidade) aos restantes mundos. Podemos ver que existem frases com exatamente a mesma intensão que (1), mas com constituintes diferentes a nível semântico. Para isso, consideremos esta frase que resulta de substituir 'gato' e 'tapete', respetivamente, pelas expressões parecidas 'gato*' e 'tapete*':

(2) Um gato* está num tapete*.

Em vez de uma função de mundos para conjuntos de gatos, ‘gato*’ tem a função ϕ_G, assim definida:

$\phi_G(w)$ = {x : x é uma cereja} se e só se, em w, (1) existe pelo menos um gato que está num tapete e existe uma cereja que está numa árvore, ou (2) não existe qualquer gato que está num tapete nem qualquer cereja que está numa árvore,

$\phi_G(w)$ = {x : x é um gato} se e só se, em w, existe um gato que está num tapete e nenhuma cereja está numa árvore.

E, de um modo semelhante, em vez de uma função de mundos para conjuntos de tapetes, a intensão de ‘tapete*’ é a função ϕ_T tal que:

$\phi_T(w)$ = {x : x é uma árvore} se e só se, em w, um gato está num tapete e uma cereja está numa árvore,

$\phi_T(w)$ = {x : x é um tapete} se e só se em w, um gato está num tapete e nenhuma cereja está numa árvore,

$\phi_T(w)$ = {x : x é um quark} se e só se em w, nenhum gato está num tapete e nenhuma cereja está numa árvore.

As intensões para ‘gato*’ e ‘tapete*’ foram pensadas de modo a que o valor de verdade de (1), no mundo atual e nos mundos meramente possíveis, seja exatamente o mesmo. A intensão de (2) é, assim como a de (1), a função que liga os mundos em que um gato está num tapete ao valor 1, e os restantes a 0. A conclusão que daqui devemos tirar é que se a única maneira de determinar os valores semânticos de ‘gato’ e ‘tapete’ forem os valores de verdade de (1) nos vários mundos possíveis, estamos igualmente autorizados a atribuir a ‘gato’ e ‘tapete’ as intensões que intuitivamente nos parecem as adequadas como, respetivamente, as funções ϕ_G e ϕ_T, ou quaisquer outras que consigamos construir de um modo análogo. É por isso adequado interpretar ‘gato’ como tendo cerejas como referentes em mundos possíveis como o atual, e, igualmente, interpretar ‘tapete’

como referindo árvores, em certos mundos, ou quarks, noutros mundos. Esta é a estranheza dos resultados de H. Putnam.

Este é um exemplo extremamente simplificado e, por isso mesmo, insuficiente para provar que a indeterminação radical afeta a linguagem em geral. A única restrição imposta foi a manutenção da intensão de uma única frase. Nada nos garante, ainda, que haja interpretações alternativas de 'gato' e 'tapete' que permitem manter intacta a intensão de todas as frases em que 'gato' e 'tapete' ocorrem. Também não foi provado que o problema se aplica a outras expressões de uma linguagem L.

Por isso, os resultados anteriores têm de ser estendidos. Uma maneira de fazer isso é mostrar que, dentro de certos limites, cada interpretação estranha de uma classe de expressões de L pode gerar os valores de verdade (atuais e meramente possíveis) corretos para as frases de L através de um reajustamento na interpretação nas restantes expressões de L. H. Putnam mostra que uma ideia próxima desta é verdadeira ao provar o seguinte teorema: «Let *L* be a language with predicates F_1, F_2, …, F_k (not necessarily monadic). Let *I* be an interpretation, in the sense of an assignment of an intension to every predicate of *L*. Then if *I* is non-trivial in the sense that at least one predicate has an extension which is neither empty nor universal in at least one possible world, there exists a second interpretation *J* which disagrees with *I*, but which makes the same sentences true in every possible world as *I* does.» (Putnam 1981: 217)

Vou apresentar aqui uma prova próxima, mas não exatamente idêntica, àquela que é apresentada por H. Putnam. Seja F_i um qualquer predicado de L que, de acordo com a interpretação I, tem como intensão a função ϕ_{Ii}, a qual lhe atribui o conjunto R_{ij} como extensão no mundo w_j. U_j é o conjunto de indivíduos possíveis existentes em w_j. Seja f_j, por sua vez, uma qualquer função bijetiva que atribui a cada membro de U_j e a cada *n*-tuplo ordenado de membros desse conjunto, respetivamente, um outro membro do mesmo conjunto ou *n*-tuplo ordenado desses indivíduos. f_j tem as seguintes propriedades: (1) para todo o $x \in U_j$ e $y \in U_j$, …, $f_j(x) \neq x$, $f_j(x, y) \neq (x, y)$, …, (2) $f_j(x, y) = (f_j(x), f_j(y))$, e (3) $f_j(x) \in U_j$, $f_j(y) \in U_j$, …. Através de f_i e R_{ij} obtemos o conjunto $f_j(R_{ij})$, sendo que $f_j(R_{ij}) \neq R_{ij}$, pelo menos nos casos em que R_i não é o conjunto vazio nem o conjunto universal. J é a interpretação que atribui a F_i a intensão ϕ_{Ji} tal que $\phi_{Ji}(w_j) = f_j(R_{ij})$.

Podemos verificar que, em cada mundo possível, as frases de L recebem exatamente o mesmo valor de verdade de acordo com I e J. Tendo em conta que a interpretação J resulta de uma aplicação uniforme de *f* sobre as extensões dos predicados de L, podemos concluir que (1) para todo o *x* e qualquer extensão R, se $x \in$ R, então $f(x) \in f(\text{R})$. Isto é o mesmo que dizer que o papel desempenhado por um objeto, de acordo com I, nas extensões dos vários predicados de L é desempenhado por um outro objeto, de acordo com J. Além disso, tendo em conta que *f* é bijetiva, (2) para qualquer R, o número de objetos que pertencem a R é o mesmo que o número de objetos que pertencem a *f*(R). Facilmente percebemos que (1) e (2) garantem a equivalência entre I e J no que diz respeito às intensões das frases de L, se L for uma linguagem de primeira ordem (mesmo com identidade). No caso de frases com quantificadores, conectivos verofuncionais e o sinal de identidade, o seu valor de verdade é determinado unicamente por propriedades e relações entre as extensões dos predicados de L que são insensíveis aos objetos que são membros dessas extensões. Os objetos têm um papel a desempenhar nas extensões, mas esse papel pode ser desempenhado por qualquer outro objeto. Se L contiver nomes, J vai ter de atribuir-lhes referentes aplicando a função *f* àqueles que eram atribuídos por I. Assim, se a referência do nome '*a*' em *w* é *a* de acordo com *I*, passa a ser *f*(*a*) de acordo com J.

Por fim, se J atribui a cada predicado F – e a cada nome, de um modo análogo – de L uma intensão que consiste na função ϕ_J tal que, para cada mundo *w*, $\phi_J(w) = f(\phi_I(w))$, sendo ϕ_I a intensão atribuída a F por I, temos então duas interpretações que (no caso de L conter pelo menos um predicado cuja extensão, em pelo menos um mundo possível, não é vazia ou universal) coincidem no que diz respeito ao valor de verdade das frases em todos os mundos, mas diferem – possivelmente de um modo radical – quanto à extensão dos predicados e à referência dos nomes.[53]

---

53 O problema pode também ser contemplado através do método de definição dos termos teóricos de D. Lewis. Podemos pensar em L como uma teoria e as expressões que nela ocorrem – deixando de lado o vocabulário lógico – como termos teóricos. Temos, então, o postulado de L, que consiste na conjunção das frases de L que consideramos verdadeiras:

$L(l_1, \ldots, l_n)$.

(Os predicados de L podem, por conveniência, ser tratados como nomes.)

Agora, pelos mecanismos apresentados no anexo 1, obtemos uma definição para cada um dos termos $l_1, \ldots, l_n$:

Consideremos agora o problema análogo que surge para a interpretação radical. Recapitulemos, antes de mais, os princípios que funcionam como restrições à interpretação radical:

(1) Se, imediatamente antes de *t*, X estava num sistema de crenças expresso pela função C, e se, em *t*, X sofre o impacto de estímulos percetivos que são evidência de que E, o seu sistema de crenças passa, em *t*, a ser aquele que é expresso por C(-/E).

(2) Se, em *t*, o sistema de crenças e desejos de X for expresso pelo par de funções C(-/E) e V, o comportamento de X em *t* deve ser aquele que, de entre o conjunto exaustivo de alternativas mutuamente exclusivas de comportamento disponíveis para X em *t*, recebe a maior utilidade esperada calculada a partir de C(-/E) e V,

Relativamente a (1) e (2), pares de funções C e V são equivalentes se atribuírem a maior utilidade esperada ao mesmo comportamento quando se dá o impacto da mesma evidência num sujeito. Não havendo mais princípios a regularem a interpretação radical, é adequado descrever as crenças e desejos tanto através das funções C e V como através das equivalentes C* e V*. Existem inúmeros pares equivalentes de funções (*ver* Lewis 1983b: 374-75) e, por isso, estamos perante uma indeterminação radical do conteúdo mental.

Seja E uma proposição relativa à evidência com que X é confrontado num dado momento, B o comportamento de X nesse momento, K um conjunto exaustivo de hipóteses de dependência, e *k* uma variável que recebe como valores possíveis os elementos de K. Como vimos antes, calcula-se a assim utilidade esperada de B, de acordo com D. Lewis:

$l_1 = \iota x_1 : \exists_1 x_2, \ldots, \exists_1 x_n \, L(x_2, \ldots, x_n)$,

…

$l_n = \iota x_n : \exists_1 x_1, \ldots, \exists_1 x_{n-1} \, L(x_1, \ldots, x_{n-1})$.

Se os valores de verdade das frases de L são capazes de determinar a referência dos termos que nelas ocorrem, então a interpretação desses termos deve ser dada pelas descrições definidas obtidas por este método – ou, dito de outro modo, cada termo deve nomear aquilo, seja o que for, que torna verdadeiras as frases que já antes determinamos serem verdadeiras.

Como todos os termos $l_1, \ldots, l_n$ teriam de ser definidos em simultâneo e, além destes, L só contém vocabulário lógico, é plausível presumir que estas são descrições impróprias: inúmeras sequências de objetos podem ser interpretações adequadas. (A apresentação do problema mais ou menos nestes termos encontra-se em (Lewis 1983b: 270 e 1984: 222-24).)

$$U(B) = \sum_k C(k/E)\ V(B \cap k)$$

Para calcular $V(B \cap k)$ só são relevantes os valores que as funções C e V atribuem aos mundos que pertencem simultaneamente a E, a B, e aos membros de K. Temos, então, para cada *k*, um conjunto de mundos relevantes para calcular $V(B \cap k)$, que é a intersecção de E, B e *k*. Podemos também ter, em vez disso, um conjunto mais alargado que consiste apenas na intersecção de E com C, que irá incluir mundos relevantes de todos os *k*. Neste conjunto, chamemos-lhe R, definimos uma função bijetiva *f*, que vai ligar cada mundo de R a um outro mundo em R. A função *f* tem de ser tal que, se, para qualquer *k*, $w \in R \cap k$, então $f(w) \in R \cap k$. A partir de *f*, de C, e de V, obtemos duas novas funções, C* e V*, tais que, para qualquer *x*, se, para qualquer $w \in R$, $C(w) = x$, vai acontecer que $C^*(f(w)) = x$; e o mesmo vai acontecer entre V e V*. Resumindo, obtemos duas funções que redistribuem uniformemente entre os mundos de R os valores que eram atribuídos por C e V, preservando, no entanto, a soma dos valores dos mundos que estão incluídos em cada *k*. Dada esta redistribuição uniforme, os pares (C, V) e (C*, V*) vão ter como resultado, para cada *k*, o mesmo valor para $(B \cap k)$. Uma função semelhante a *f* pode ser aplicada à intersecção de E com qualquer proposição que pertence ao conjunto A, e essas modificações podem ser capturadas no par (C*, V*). Como se trata de um conjunto de proposições mutuamente exclusivas, nenhuma proposição nele contida tem um mundo em comum. Assim, a redistribuição de valores atribuídos aos mundos em cada uma das proposições de A não afeta de modo algum as restantes. No final do processo, a utilidade esperada atribuída a cada uma das proposições de A, e não apenas a B, continua intacta quando calculada tanto por (C, V) como por (C*, V*).

Pela restrição que foi imposta à função *f*, deixa-se intacto em C* o valor atribuído a cada uma das hipóteses de dependência por C. Podemos expandir a indeterminação, modificando estes valores, mantendo a utilidade de cada comportamento idêntica, do seguinte modo. Para cada conjunto obtido pela intersecção de E com cada um dos elementos de A (chamemos-lhes R_1, …, R_n, tendo A *n* elementos) especifica-se subconjuntos dos quais fazem parte apenas os mundos contidos numa certa proposição que pertence a K. Se K tem *m* membros, teremos os conjuntos, assumindo uma qualquer

ordenação de K, $R_1 \cap k_1$, …, $R_1 \cap k_m$, …, $R_n \cap k_1$, …, $R_n \cap k_m$. Agora ordenamos cada um desses subconjuntos, tornando-os em sequências, que serão aqui designadas pelo par ordenado dos números indexados ao correspondente conjunto R e proposição *k* (uma delas vai ser chamada (*n*, *m*), por exemplo). Definimos também uma função bijetiva f_1, que liga cada *k* a outro *k*. Agora, faz-se a seguinte alteração entre C* e uma nova função C** (e, também, entre V* e V**): para qualquer *x*, se w_1 for o membro *i* da sequência (*n*, *m*), $C^*(w_1) = x$, $f(k_m) = k_j$, e w_2 for o membro *i* da sequência (*n*, *j*), então $C^{**}(w_2) = x$. Se estas alterações forem aplicadas uniformemente, teremos, no final do processo, de acordo com C** e V**, exatamente os mesmos valores para a utilidade de cada comportamento que tínhamos de acordo com (C, V) e (C*, V*), mas diferentes probabilidades para as várias hipóteses de dependência relativamente àquelas que eram atribuídas por C e C*.[54]

De acordo com D. Lewis, «Putnam's thesis is incredible. We are in the presence of a paradox […]. It is out of the question to follow the argument where it leads. We know in advance that there is something wrong, and the challenge is to find out where.» (Lewis 1985: 221) Ao contrário de H. Putnam, que olha para o problema com que se confronta mais como uma premissa de um argumento do que como uma dificuldade a ser ultrapassada,[55] D. Lewis encara-o como demonstrando a inadequação de assumir um

---

54 Estou aqui a assumir que os mundos que fazem parte de cada uma das hipóteses de dependência fazem parte de uma das proposições de alternativas de comportamento de X. A minha ideia é a seguinte. Para ser verdade que se X faz A num certo momento, os eventos B, C, D, …, dependerão da sua ação, e que, ao invés, E, F, G, …, já não dependerão, o mundo tem de ser tal que X possa fazer A nesse momento. Se assim for, a minha suposição é correta.

55 H. Putnam pretende usar este problema como uma premissa para atacar a posição a que chama *realismo metafísico*, que consiste em afirmar que «the world consists of some fixed totality of mind-independent objects. There is exactly one true and complete description of 'the way the world is'. Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things.» (Putnam 1981: 49) Em "Realism and Reason" (1977), nota que uma das principais consequências da posição realista é a de que «*truth* is supposed to be *radically non-epistemic* – […] and so the theory that is "ideal" from the point of view of operational utility, inner beauty and elegance, "plausibility", simplicity, "conservatism", etc., *might be false*. "Verified" (in any operational sense) does not imply "true", on the metaphysical realist Picture, even in the ideal limit.» (Putnam 1977: 485)

Em (Putnam 1977: 486), parece ser sugerido que o problema da indeterminação radical coloca em causa o realismo metafísico do seguinte modo: se uma teoria impecável do ponto de vista da aceitabilidade puder ser falsa – como o realismo metafísico implica –, então existe uma relação, entre os termos da teoria e as coisas no mundo, que determina que algumas frases da teoria são falsas. No entanto, se as restrições que temos em mãos para interpretar os termos da teoria, i. e. atribuir-lhes uma certa relação com as coisas no mundo, forem exatamente aquelas que governam a aceitabilidade racional da teoria, nenhuma das interpretações possíveis (de entre o vasto número dessas interpretações) pode tornar a teoria falsa. A solução passaria, claro, por introduzir mais restrições à interpretação dos termos que sejam independentes dos princípios de aceitabilidade da teoria. Mas esta não parece ser uma posição que H. Putnam esteja pronto a

número escasso de restrições para a interpretaeção radical. Neste sentido, D. Lewis tem de admitir que a psicologia popular tem muito mais a dizer sobre a racionalidade dos sujeitos. Estivemos apenas a considerar princípios acerca da racionalidade instrumental e, segundo ele, «instrumental rationality, though it is the department of rationality that has proved most tractable to systematic theory, remains only one department among others. We think that some sorts of belief and desire (or, of dispositions to believe and desire in response to evidence) would be unreasonable in a strong sense – not just unduly skeptical or rash or inequitable or dogmatic or wicked or one-sided or short-sighted, but utterly unintelligible and nonsensical.» (Lewis 1986b: 38)

3.4 – A naturalidade das propriedades

Assim, a psicologia popular não nos diz apenas que normalmente os agentes escolhem aquele curso de ação que melhor serve os seus desejos, de acordo com as suas crenças, e que mudam as suas crenças conforme a evidência com que vão sendo confrontados, mas distingue também entre a razoabilidade e inteligibilidade dessas crenças e desejos. Os princípios que tocam nestas propriedades do conteúdo das atitudes são classificados por D. Lewis como princípios da *caridade* ou *humanidade*, e são assim descritos:

> «If we rely on principles of fit to do the whole job, we can expect radical indeterminacy of interpretation. We need further constraints, of the sort called principles of (sophisticated) charity, or of 'humanity'. Such principles call for interpretations according to which the subject has attitudes that we would deem reasonable for one who has lived the life that he has lived. […] These principles select among conflicting interpretations that equally well conform to the principles of fit. They impose *a priori* – albeit defeasible – presumptions about what sorts of things are apt to be believed and desired; or rather, about what dispositions to develop beliefs and desires, what inductive biases and basic values, some may rightly be interpreted to have.» (Lewis 1983b: 375)

aceitar. «So what *further* contraints on reference are there that could single out some other interpretation [i. e., uma interpretação que torne falsa a teoria] as (uniquely) "intended" […]?», pergunta H. Putnam, acrescentando de seguida: «The supposition that even an "ideal" theory (from a pragmatic point of view) might *really* false appears to collapse into *uninteligibility*» (Putnam 1977: 486).

Podemos então propor uma primeira formulação do princípio de caridade, próxima daquela que é apresentada na passagem que acabei de citar e em (Lewis 1974: 336):

(PC) As pessoas acreditam e desejam normalmente aquilo que *nós* – deixando esta expressão talvez propositadamente ambígua – consideramos ser adequado acreditar e desejar dada a evidência que têm ao seu dispor e a aprendizagem que receberam.

Repare-se que, de acordo com (PC), não se consideram os outros sujeitos como tendo as mesmas crenças e desejos que nós. Podemos pensar, inteligivelmente, que os outros têm crenças falsas ou desejos que colidem com os nossos. A caridade interpretativa imposta por (PC) deve ser compreendida como dizendo respeito não diretamente à verdade, mas à adequação racional das crenças. Se X viveu em condições nas quais não teve acesso a evidência adequada ou aquela com que foi confrontado era de algum modo enganosa, é mais caridoso interpretar X como acreditando em proposições falsas, do que atribuir-lhe crenças verdadeiras para as quais não está de modo algum justificado.[56]

Este princípio, pelo menos assim apresentado, parece deixar a correção de uma interpretação completamente relativa àquilo que são as ideias do(s) intérprete(s). Em (Lewis 1983b: 375), este princípio é expandido de um modo que anula, pelo menos

[56] Apesar de não impor uma aproximação entre as nossas atitudes e as atitudes de X, D. Lewis diz-nos que: «there must exist some common inductive method which would lead to approximately our present systems of belief if given our life histories of evidence, and which would likewise lead to approximately the present system of beliefs ascribed to Karl […] if given Karl's life history of evidence»; e, acrescenta: «As for desires, there must exist some common underlying system of basic intrinsic values which would yield approximately our systems of desires if given our systems of beliefs, and which would likewise yield approximately the system of desires ascribed to Karl […], if given the system of beliefs ascribed to Karl.» (Lewis 1974: 336) Esta proposta diz-nos, usando as ferramentas conceptuais antes apresentadas, que a função V deve ser idêntica para todos os sujeitos – devemos projetar nos outros os mesmos desejos básicos que temos – sendo que a diferença, para cada proposição P entre a nossa V(P) e a V(P) de X reside no valor de C(P/E), sendo E a evidência disponível. A função C(-/E), por sua vez, deve partir em todos os sujeitos de uma mesma distribuição *a priori* de probabilidades C. Neste caso, não devemos interpretar C como as crenças que um sujeito tem num certo instante de tempo, mas nas disposições doxásticas que um sujeito mantém durante a sua vida. Assim, a diferença entre as nossas crenças e as crenças de X dependem do total de evidência com a qual fomos confrontados no presente e no passado. Assim, apesar de (PC) não obrigar os vários sujeitos a terem as mesmas crenças e os mesmos desejos, sugere tendencialmente uma aproximação entre as disposições e os valores básicos de todos eles.

parcialmente, essa relatividade. Em vez de se falar apenas naquilo que *nós* consideramos adequado ou razoável, é estabelecida uma preferência objetiva pelas interpretações que envolvem as propriedades mais *naturais*, aquelas que D. Lewis classifica como elegíveis. Temos, então, que:

(PC*) Além de (PC), o conteúdo das atitudes das pessoas faz parte de uma classe de proposições (ou propriedades) elegíveis.

Como se viu anteriormente (*ver* secção 1.5.5), D. Lewis propõe uma teoria das propriedades extremamente simples e clara, construída a partir do seu realismo modal. De acordo com essa teoria, uma propriedade é um conjunto de objetos possíveis. A propriedade de ser um morcego, por exemplo, é o conjunto cujos elementos são os morcegos espalhados por todo o espaço lógico – que habitam o mundo atual ou qualquer um dos restantes mundos possíveis. A propriedade de ser um morcego ou um rato, por sua vez, é o conjunto dos morcegos e dos ratos atuais e possíveis. O mesmo se aplica às relações, com a única diferença de estas serem identificadas com conjuntos de pares, triplos ordenados, …, de objetos possíveis. A relação expressa pelo predicado '*x* ama *y*', por exemplo, é o conjunto de pares ordenados de pessoas possíveis (*x*, *y*) tais que *x* ama *y* (Lewis 1983b, 343-44, 1986b: 50-2).

De acordo com esta teoria, é verdade que cada propriedade F, expressa pelo predicado 'F', é idêntica ao conjunto $\{x : F(x)\}$(se permitirmos que o domínio de *x* inclua *todos* os objetos possíveis). Mas, visto que D. Lewis não estabelece qualquer restrição relativamente a que conjuntos são propriedades, não é apenas verdade que qualquer propriedade é idêntica ao conjunto das suas instâncias, mas também que qualquer conjunto é uma propriedade. Assim, para quaisquer objetos aleatoriamente selecionados – ou pares, ou triplos ordenados, …, arbitrariamente escolhidos – temos uma propriedade – ou relação – que é instanciada apenas por eles (Lewis 1983b: 346).

Deste modo, existem muito mais propriedades do que aquelas que são expressas pelos nomes comuns das linguagens humanas, como 'morcego', 'rato', 'dor', 'eletrão', e por aí em diante. O seu número ultrapassa também largamente o daquelas que podem ser,

com maior ou menor esforço, especificadas a partir da aplicação de operações booleanas sobre os predicados que temos à disposição, como 'morcego ou rato', 'morcego e rato', 'morcego, no caso de estar a chover e rato, no caso de não estar a chover', e por aí em diante. Esta situação traz consigo algumas consequências relevantes. «Because properties are so abundant», diz-nos D. Lewis, «they are undiscriminating. Any two things share infinitely many properties, and fail to share infinitely many others. That is so whether the two things are perfect duplicates or utterly dissimilar». Conclui, assim, que «properties do nothing to capture facts of resemblance.» (Lewis 1983b: 346)

É por esta razão que D. Lewis pensa que «an *adequate* theory of properties is one that recognises an objective difference between natural and unnatural properties.» (Lewis 1983b: 347) As propriedades – no sentido lato – são incapazes de nos fornecer uma descrição satisfatória do carácter qualitativo dos objetos e dos vários mundos; o corte que cada propriedade efetua na realidade é insensível às semelhanças e às diferenças entre os objetos. Para estes propósitos, D. Lewis considera imprescindível uma distinção entre propriedades que são, e aquelas que não são, *naturais*. (No anexo 3 encontra-se uma apresentação das várias teorias que D. Lewis considera adequadas para a explicar esta distinção.) As propriedades naturais, como D. Lewis as descreve, «are intrinsic, they are highly specific, the sets of their instances are *ipso facto* not entirely miscellaneous, there are only just enough of them to characterise things completely and without redundancy.» (Lewis 1986b: 60) A abundância das propriedades compreendidas como quaisquer conjuntos de *possibilia* já não é replicada pelas propriedades naturais. A escassez destas últimas será extremamente importante para que a solução de D. Lewis ao problema da indeterminação radical do conteúdo seja viável.

Mesmo entre as propriedades consideradas naturais, D. Lewis pensa que é adequado admitir uma hierarquia (Lewis 1983b: 347, 1986b: 61). Apenas algumas destas são propriedades *perfeitamente* naturais: aquelas propriedades das quais, de acordo com D. Lewis, se pode dizer *a priori* que qualquer diferença no carácter qualitativo de dois mundos exige alguma diferença no seu padrão de instanciação nesses mundos (Lewis 1992a: 218, 1994: 412-13). Equivalentemente, acontece que se dois mundos forem idênticos ao nível da instanciação das propriedades perfeitamente naturais, são idênticos

simpliciter – são semelhantes em todos os aspetos. É deste modo que se pode dizer que estas propriedades são as necessárias para descrever o mundo de um modo completo e sem redundância.

Os casos paradigmáticos de propriedades perfeitamente naturais são as propriedades mais ou menos próximas daquelas que aparecem referidas na teoria física contemporânea, como massa e carga. Noutros mundos possíveis, talvez outras propriedades muito diferentes daquelas conhecidas pela física são instanciadas. Mas, pelo menos no mundo atual, D. Lewis acredita que o materialismo é verdadeiro – ou seja, a descrição que a física faz do mundo é uma descrição completa. D. Lewis considera que a melhor forma de expressar esta ideia usa a noção de propriedade naturais do seguinte modo:

(M) De entre os mundos em que nenhuma propriedade perfeitamente natural *estranha* ao mundo atual é instanciada, quaisquer desses mundos que diferem em algum aspeto, diferem também no padrão de instanciação de propriedades físicas.[57]

---

57 Esta formulação é um dos casos que D. Lewis pretende apresentar como uma vantagem teórica da distinção entre propriedades naturais e não naturais. Repare-se que haveria alguma dificuldade em expressar a ideia por trás da tese materialista sem recorrer à noção de propriedades perfeitamente naturais. Por exemplo, considere-se esta formulação:

(M*) Quaisquer mundos que diferem em alguns respeito, diferem quanto à instanciação de propriedades físicas.

Ao contrário de (M), (M*) tornaria o materialismo verdadeiro em todos os mundos possíveis – uma verdade necessária, portanto (Lewis 1983b: 362). Mas, mesmo que o materialismo seja verdadeiro, é-o contingentemente.

Uma estratégia insatisfatória, ainda que resulte numa formulação verdadeira, consiste em isolar os mundos em que o materialismo é verdadeiro:

(M**) Em nenhuns mundos em que o materialismo é verdadeiro, pode haver uma diferença em algum respeito sem haver uma diferença física.

Neste caso, (M**) pressupõe já a diferença entre os mundos em que o materialismo se aplica e aqueles em que tal não acontece. Além do mais, se quisermos dizer que a classe de mundos materialistas é aquela em que não há diferença sem diferença física, temos o problema de que, como D. Lewis afirma, «[…] there are many such classes. In fact any world, however spirit-ridden, belongs to such a class.» (Lewis 1983b: 363)

Por fim, também não seria adequado formular o materialismo deste modo:

(M***) Não pode haver qualquer diferença em dois mundos que obedecem às mesmas leis da natureza do mundo atual sem que haja entre eles uma diferença a nível físico.

A referência às leis da natureza é insatisfatória, porque podem existir mundos onde existem objetos não físicos, mas onde se verificam as mesmas leis da natureza que no nosso mundo; e podem existir também mundos puramente físicos em que as leis da natureza não excluem a existência de objetos não físicos (Lewis 1983b: 363).

(Lewis 1983b: 364)

Repare-se que (M) é uma aplicação particular do princípio geral antes mencionado, de acordo com o qual a diferença entre mundos depende de uma diferença nas propriedades naturais. Acontece que em (M), assume-se que as propriedades perfeitamente naturais instanciadas no mundo atual são apenas propriedades físicas.

As propriedades perfeitamente naturais, além de descreverem os mundos de um modo completo, servem também para descrever totalmente o carácter qualitativo intrínseco de qualquer objeto. Propriedades *intrínsecas* – contrariamente às *extrínsecas* – são aquelas que um objeto instancia independentemente das relações que estabelece com o que lhe rodeia; compare-se, por exemplo, a propriedade de estar localizado nos Montes Apalaches com a propriedade de ser metálico. D. Lewis propõe que analisemos as propriedades intrínsecas através da noção de *duplicado*, cuja definição, por sua vez, envolve uma referência às propriedades perfeitamente naturais:

(D) Dois objetos possíveis[58] são duplicados se e só se partilham todas as suas propriedades perfeitamente naturais.[59]

(Lewis 1983b: 356, 1986b: 61-2)

Indiretamente, (M***) – mesmo que fosse adequada – não seria uma formulação independente da noção de propriedades naturais, pelo facto de D. Lewis considerar que precisamos de apelar a estas propriedades para analisar a noção de leis da natureza (Lewis 1983b: 365-68).

58 As relações também devem ser, de acordo com D. Lewis, distinguidas de acordo com a sua naturalidade (Lewis 1986b: 61). A partir das relações perfeitamente naturais, podemos definir uma relação de duplicação entre pares (ou triplos ordenados, …) de objetos. Dois pares (*a*, *b*) e (*a**, *b**) são duplicados se e só se (1) *a* e *a**, assim como *b* e *b**, têm as mesmas propriedades perfeitamente naturais, e (2) as relações perfeitamente naturais estabelecidas entre *a* e *b* são as mesmas que aquelas que são estabelecidas entre *a** e *b** (Lewis 1983b: 356).

59 Talvez seja necessária mais alguma complexidade. D. Lewis admite que (D) é correta se aceitarmos como perfeitamente naturais as propriedades *estruturais*, «properties having to do with the way a thing is composed of parts with their own properties and relations.» (Lewis 1986a: 62) Vamos supor, por exemplo, que ser um eletrão, um protão e um neutrão são propriedades perfeitamente naturais; e que também é natural a relação que se tem de estabelecer entre eles para que formem um átomo. Podemos então admitir como natural a propriedade de ser um átomo – que consiste na propriedade de todas as estruturas cujas partes instanciam certas propriedades naturais e estabelecem entre si uma relação igualmente natural.

Se não o pretendermos admitir, temos de reformular (D) do seguinte modo:

(D*) Dois objetos possíveis são duplicados se e só se (i) têm as mesmas propriedades naturais e (ii) as suas partes podem ser colocadas em correspondência de tal modo que partes correspondentes têm as mesmas propriedades naturais e estabelecem as mesmas relações naturais (Lewis 1986b: 61).

Agora, podemos dizer que as propriedades intrínsecas são aquelas que, se instanciadas por um objeto, têm de ser também instanciadas por qualquer um dos seus duplicados (Lewis 1986b: 62). Dito de outro modo, as propriedades intrínsecas são supervenientes a partir das propriedades perfeitamente naturais do seguinte modo: nenhuns objetos possíveis podem diferir nas suas propriedades intrínsecas sem que haja alguma diferença nas suas propriedades perfeitamente naturais. [60] Por exemplo, se admitirmos como perfeitamente naturais as propriedades de ser um átomo de hidrogénio e um átomo de oxigénio e a relação R que tem de se estabelecer entre dois átomos de hidrogénio e um de oxigénio para formarem uma molécula de água, dizemos que se um objeto tiver a propriedade de ser composto por dois átomos de hidrogénio e um de carbono que entre si estabelecem a relação R, esse objeto tem também a propriedade de ser uma molécula de água; e, ainda, dizemos também que qualquer duplicado desse objeto é também uma molécula de água – o que significa que esta é uma propriedade intrínseca.

Além destas propriedades cuja naturalidade é máxima, D. Lewis admite ainda que se considerem como naturais algumas propriedades que, mesmo que ligeiramente disjuntivas e extrínsecas, devem ser assim consideradas pelo facto de poderem ser definidas de um modo não muito complicado a partir das propriedades perfeitamente naturais (Lewis 1986b: 61). De entre estas pode haver uma gradação entre as mais e menos naturais, mas D. Lewis não apresenta nenhum critério para determinar o grau de naturalidade de cada proposição.

Ao fornecer-nos uma teoria das propriedades, D. Lewis oferece-nos também uma teoria das proposições. Algumas das propriedades são proposições – mais concretamente, aquelas que são instanciadas apenas por mundos possíveis inteiros, em vez das suas

---

60 A noção de duplicado pode também ser aplicada a mundos: quaisquer mundos idênticos no padrão de instanciação de propriedades naturais são duplicados. O princípio segundo o qual diferenças entre mundos implicam diferenças a nível das propriedades naturais pode ser capturada assim:

(1) Nenhuns mundos que diferem em algum respeito são duplicados.

Ou, equivalentemente:

(1') Quaisquer mundos que são duplicados são semelhantes em todos os aspetos.

Com isto, podemos, por exemplo, reformular (M) do seguinte modo:

(M') Quaisquer mundos em que não ocorram propriedades estranhas ao mundo atual e que exemplifiquem o mesmo padrão de instanciação de propriedades físicas são duplicados.

partes. E, «sets of worlds», reconhece D. Lewis, «may accordingly be divided into the more and less natural. This is automatic, given the division of properties plus the identification of propositions with properties of worlds.» (Lewis 1986a: 61) Algumas proposições capturam maiores semelhanças entre os mundos em que são verdadeiras do que outras. Isso é facilmente percetível na comparação entre as proposições disjuntivas e os seus disjuntos. Que existem gatos e que existem cerejas são proposições certamente mais naturais do que a proposição de acordo com a qual existem gatos ou existem cerejas. Presumivelmente, as proposições mais naturais são aquelas que distinguem os mundos de acordo com o seu padrão de instanciação de propriedades perfeitamente naturais – e, do mesmo modo, as restantes proposições são mais ou menos naturais de acordo com a naturalidade das propriedades através das quais distinguem o mundo.[61]

Em (Lewis 1983b), apesar de D. Lewis apresentar esta distinção entre as propriedades como parte da solução adequada para o problema da indeterminação radical, não nos apresenta em detalhe o modo como essa estratégia deve funcionar. É conveniente considerar como é que a naturalidade das propriedades permite, em particular, distinguir entre pares de funções interpretativamente equivalentes.

Creio que provavelmente não podemos aplicar uma distinção baseada na naturalidade diretamente às próprias funções consideradas. Vamos assumir que um determinado valor de probabilidade atribuído a um mundo por C e um valor de desejabilidade atribuído por V expressam a indiferença de um sujeito X (designemos estes valores, respetivamente, por i_c e i_v). Agora, estipulamos que é verdade que:

(1) X acredita que P se e só se $C(P) > i_c$.

(2) X deseja que P se e só se $V(P) > i_v$.

Para garantir a aplicação do princípio da bivalência neste caso, dizemos também que:

61 Nesta conceção de proposições, dizemos que as proposições distinguem os mundos de acordo com o padrão de instanciação de propriedades e não que as proposições incluem propriedades. Frases contêm predicados que podem ter como valor semântico propriedades – ou intensões construídas de acordo com propriedades –, mas proposições não têm uma estrutura sintática como as frases, pelo menos se forem compreendidas como conjuntos de mundo possíveis.

(1*) X não acredita que P se e só se $C(P) \leq i_c$.

(2*) X não deseja que P se e só se $C(P) \leq i_v$.[62]

Se pretendermos aplicar o princípio de caridade que estabelece uma preferência pela atribuição de proposições mais naturais como conteúdo das atitudes de um sujeito, uma estratégia plausível consistiria em dizer que a escolha entre os pares de funções equivalentes faz-se tendo em conta qual dos pares deixa como verdadeira mais frases que afirmam que X acredita (ou deseja) proposições mais naturais. O problema desta estratégia é a de que a ausência de crença na proposição P implica a crença na proposição ~P, se não houver indiferença relativamente a P. Ora, para cada frase de acordo com a qual X não acredita numa proposição extremamente pouco natural, é verdadeira uma frase que nos diz que X acredita na sua negação, que é, se não mais, pelo menos tão pouco natural como a primeira.

A outra estratégia que vislumbro é a de preferir o par de funções que deixa o sujeito indiferente ao maior número de proposição inadmissivelmente não naturais. Mas esta estratégia também não funciona. Repare-se que, para quaisquer proposições P e Q tais que $P \subset Q$, $C(P) \leq C(Q)$, o que significa que se $C(P) > i_c$, então $C(Q) > i_c$. Mas Q pode ser uma proposição extremamente menos natural do que P. Q pode ter sido obtida aleatoriamente através da união da proposição natural P com outras proposições sem ter em conta a semelhança entre os mundos. Assim, não é viável dizer-se que uma função C pode expressar a crença em proposições naturais, expressando ao mesmo tempo indiferença relativamente a um grande número de proposições pouco naturais.

Uma outra estratégia passa por especificar alguns valores atribuídos por C e V como aqueles que expressam realisticamente as crenças e desejos de X. Por exemplo, para C esses valores podem ser 0, 0.25, 0.5, 0.75 e 1. Para distinguir os vários pares de funções, diríamos que a mais adequada é aquela que atribui esses valores um menor número de proposições pouco naturais. Talvez esta seja uma estratégia viável.

---

62 De acordo com esta estipulação, 'X não acredita que P' não é equivalente a 'X acredita que ~P'. Repare-se que para que esta última frase seja verdadeira é necessário que $C(P) < i_c$, o que é o mesmo que dizer que $C(\sim P) > i_c$. Esta é uma condição mais forte do que aquela que (1*) especificam para 'X não acredita que P seja verdadeira'.

No entanto, temos em mãos uma estratégia muito mais simples, que nos deixa liberdade para não ter de escolher diretamente entre os vários pares de funções. Numa interpretação, como se viu antes, pretende-se atribuir um triplo ordenado formado por uma atitude, um grau e uma proposição (A, *g*, P) a um estado físico F, se considerarmos F como sendo idêntico a uma única atitude; ou, alternativamente, atribui-se a F um conjunto desses triplos se considerarmos F como um total sistema de crenças e desejos. Um sujeito X tem as crenças e desejos especificados por um conjunto de triplos ordenados se X estiver num estado físico que seja idêntico a um sistema de crenças e desejos cuja interpretação atribui esse conjunto, ou se X estiver em vários estados físicos cada um dos quais é interpretado por um dos triplos ordenados do conjunto.

Agora, indiretamente, podemos dizer que X tem as atitudes expressas pelo par de funções (C, V) se e só se, para cada triplo (C, *g*, P) e (D, *g*, P) que pertence ao conjunto da interpretação, acontece que C(P) = *g* ou V(P) = *g*, respetivamente. Para manter a estipulação expressa por (1)-(1*) e (2)-(2*), dizemos que $g > i_c$ ou $g > i_v$, pelo que pretendemos afirmar, para cada (A, *g*, P) atribuído a X, que X acredita (ou deseja) que P. Tendo em conta que a informação contida nas funções C e V tem de dizer respeito a *todos* os mundos possíveis e, *a fortiori*, a todas as proposições, excede em grande medida aquela que é expressa por um qualquer conjunto finito de triplos ordenados. Uma das consequências disto é a de que vários pares de funções expressam as crenças e desejos de X de um modo adequado, de acordo com a informação fornecida pelos triplos ordenados.

Podemos resolver parcialmente assim esta situação. Estipulamos que C e V devem atribuir preferencialmente valores de indiferença a cada mundo (que resulta de distribuir uniformemente o valor máximo de probabilidade pelos pontos do espaço de possibilidades) e ajustar esses valores apenas quando necessário para acomodar as proposições expressas pelos triplos ordenados. Mesmo que exista ainda indeterminação entre vários pares de funções, esta é pelo menos atenuada. C e V passam, de um certo modo, a ser descrições artificiais das atitudes de X, mas a situação é tolerável, já que o que se pretende destas funções é que nos forneçam um cálculo adequado da utilidade esperada do comportamento e das relações entre as crenças e a experiência.

Podia agora ser colocada a seguinte questão: se a interpretação não usa imediatamente as funções C e V, será que o problema da indeterminação do conteúdo não é, então, um problema fictício criado pela suposição de que a interpretação consistia exatamente na escolha entre os vários pares dessas funções? A resposta, creio eu, deve ser negativa. Vejamos que a indeterminação entre pares de funções equivalentes é transmitida a conjuntos equivalentes de triplos ordenados. Se a informação contida num conjunto de triplos ordenados é expressa por um – ou mais do que um – par de funções (C, V), a interpretação de F por um certo conjunto que é expresso por (C, V) é equivalente a um outro conjunto expresso por (C*, V*).

Ora, aplicando agora o princípio da caridade, devemos preferir a atribuição de triplos que contenham menos proposições extremamente pouco naturais. Indiretamente, ao escolhermos entre vários triplos ordenados ou conjuntos de triplos ordenados possíveis para a interpretação, estamos a eliminar alguns pares de funções.

Resolveu-se, então, o problema da indeterminação radical. Apenas as proposições mais naturais são elegíveis como conteúdo de atitudes. A elegibilidade do conteúdo, além de restringir *a priori* as crenças e desejos possíveis de um sujeito, permite também dizer que devemos descrevê-lo como considerando mais plausível, antes da evidência, certas hipóteses mais naturais do que aquelas que o são menos; desse modo, dada a evidência disponível, deve formar certas crenças em vez de outras (Lewis 1986b: 38, 1994: 427-28). É por isso, por exemplo, que a maior parte de nós acredita que as esmeraldas que veremos no futuro serão verdes, em vez de verduis.

Por fim, consideremos um problema para esta abordagem. Vimos antes que as propriedades referidas pelas teorias físicas são o exemplo paradigmático de propriedades perfeitamente naturais, pelo menos no mundo possível atual. No entanto, poucas pessoas têm atitudes relativas à instanciação de propriedades físicas no mundo – de facto, até a um certo ponto na história da nossa espécie, certamente *ninguém* tinha crenças diretamente relacionadas com essas propriedades; e, possivelmente, ainda hoje ninguém as tem, porque aquilo que a física fundamental nos revelará no futuro pode ser muito distinto daquilo que a física contemporânea tem a dizer.

Normalmente, as proposições em que os seres humanos acreditam distinguem o mundo através de propriedades um pouco mais extrínsecas – propriedades que, ainda que naturais, descrevem os objetos e o mundo a um nível superior, em termos causais, teleológicos, entre outros. Pense-se, por exemplo, nos estados mentais. Se a análise funcionalista de D. Lewis estiver correta, quando X acredita que A está a sentir dor, está a acreditar na proposição de acordo com a qual o nosso mundo é dos locais do espaço lógico em que A está num estado – físico ou não físico, seja ele qual for – que desempenha o papel causal R. Mais ainda, quando X acredita que está diante de uma cadeira, está a acreditar que vive num dos mundos em que à sua frente está um objeto – feito de um qualquer material, que ele muito provavelmente não sabe descrever com suficiente detalhe em termos físicos – que serve para as pessoas se sentarem. Outra situação é aquela em que falamos de certas propriedades pelo modo com nos afetam na experiência, como é o caso das cores. (Possivelmente, também pensamos em materiais naturais como água ou ouro através do modo como chegam aos nossos sentidos. Seja como for, a maior parte de nós não pensa neles de acordo com a sua estrutura física ou química.) No entanto, é plausível dizer que, de acordo com (PC*), devemos preferir interpretações que atribuem proposições relativas às propriedades físicas, se possível.

Fica em aberto, creio, se este é realmente um problema para a teoria da interpretação de D. Lewis. Pode, no entanto, responder-se ao mesmo dizendo que (PC) é também um princípio da psicologia popular que restringe a interpretação; e, plausivelmente, não é razoável interpretar criaturas como os seres humanos – com a evidência que têm disponível e com a aprendizagem que receberam – como tendo crenças (e desejos) relativos às propriedades físicas do mundo. Esta resposta parece-me adequada, mas aumenta uma das preocupações que antes levantei em relação à teoria de D. Lewis: teremos uma relativização na interpretação às ideias do(s) intérprete(s). Aquilo que é ou não é razoável depende em grande medida daquilo que *nós* consideramos razoável.

3.5 – A interpretação radical da linguagem

De entre o comportamento que é racionalizado pelo sistema de crenças e desejos de um sujeito X está o comportamento linguístico. Através das crenças e desejos que

envolvem este comportamento, D. Lewis sugere que é possível atribuir condições de verdade às frases usadas por X; ou, equivalentemente, podemos especificar qual é a linguagem falada por X.

De acordo com o uso desta noção em *Convention* (Lewis 1969a) e "Languages and Language" (Lewis 1975), uma *linguagem possível* L é um objeto abstrato através do qual certas frases recebem uma interpretação – mais concretamente, uma função que atribui uma proposição (ou a correspondente função de mundos possíveis para valores de verdade) a expressões que são consideradas as frases de L. Para ser mais rigoroso, exige-se uma maior complexidade, tendo em conta que algumas frases são *indexicais*. Uma linguagem deve ser capaz de atribuir diferentes interpretações a uma frase dependendo de certos fatores contextuais – paradigmaticamente, o mundo possível, o falante, o lugar e o tempo da elocução. Outros fatores relevantes podem ser, como D. Lewis nota em "General Semantics" (1970b: 24), a audiência, os objetos apontados ou o discurso antecedente.

D. Lewis propõe que se especifique implicitamente todos estes fatores através de um triplo ordenado formado por um mundo, um falante e um instante de tempo; ficando os restantes fatores a serem determinados, de uma maneira potencialmente complexa, por estes três. Visto que D. Lewis considera que um indivíduo é parte de apenas um mundo, outra hipótese que tem disponível consiste em tratar um contexto simplesmente como um falante – ou, melhor ainda, uma parte temporal de um falante (Lewis 1980a: 85-6, 1983c: 230, 1986b: 40-1). Além da indexicalidade, há ainda a considerar a possibilidade de existirem frases ambíguas. Para acomodar esta situação, uma linguagem pode atribuir várias interpretações a uma frase. Assim, resumindo, uma linguagem é uma função de pares ordenados (*e*, *c*), formados por expressões e contextos, para conjuntos de interpretações – na melhor das hipóteses, um conjunto com apenas um elemento (Lewis 1969a: 160-65, 1975: 13-4). Podemos agora dizer que:

(V) Uma frase Ψ é verdadeira-em-L no contexto *c* se e só se $L(\Psi, c) = \phi$ (ou, se Ψ for ambígua, $L(\Psi, c) = I$ e $\phi \in I$) e $\phi(@) = 1$.

Resta saber o que torna uma linguagem possível L aquela que é adequada como a interpretação das frases de um sujeito X ou de uma população P – i. e., o que torna L a linguagem atual de X ou P. A proposta de D. Lewis a esse respeito consiste em dizer que L é a linguagem usada pelos membros da população P se existir em P uma *convenção* de sinceridade e confiança em L (Lewis 1969a: 177-95, 1975: 7-12).

Antes de vermos em que consiste uma convenção de sinceridade e confiança em L é importante ter em conta como D. Lewis entende a noção de *convenção* em geral:

> «Conventions are regularities in action, or in action and belief, which are arbitrary but perpetuate themselves because they serve some sort of common interest. Past conformity breeds future conformity because it gives one a reason to go on conforming; but there is some alternative regularity which could have served instead, and could have perpetuated itself in the same way if only it had got started.» (Lewis 1975: 5)

Neste sentido, uma convenção é uma prática racional, que se mantém estável numa comunidade devido ao modo como essa prática serve os interesses dos membros da comunidade, de acordo com as suas crenças. As crenças aqui relevantes dizem respeito às expetativas de conformidade à prática mantidas mutuamente pelos vários intervenientes. Para clarificar esta ideia, D. Lewis formula, em (Lewis 1975: 5-6), as seguintes condições – que considera como necessárias e conjuntamente suficientes – para que uma regularidade R seja uma convenção numa população P: (1) todos os membros de P se conformam a R, (2) todos os membros de P acreditam que os restantes membros de P se conformam a R, (3) a crença de que os outros se conformam a R fornece a cada membro de P uma razão para se conformar também a R, (4) há uma preferência geral em P para que todos se conformem a R e que não existam algumas exceções, (5) existe pelo menos uma alternativa possível a R, R*, tal que R* cumpre as condições (3) e (4); e, por fim, (6) as condições anteriores são conhecimento comum entre os membros de P, ou seja, todos as sabem, e todos sabem que os outros as sabem, e por aí em diante.

Agora, podemos dizer que uma convenção de sinceridade e confiança em L numa população P é uma regularidade de sinceridade e confiança em L que cumpre, em P, as condições (1)-(6). Verifica-se em P uma regularidade de sinceridade em L quando os membros de P tentam não dizer frases que não sejam verdadeiras-em-L, e uma

regularidade de confiança em L se os membros de P tendem a acreditar que as frases ditas pelos outros falantes são verdadeiras-em-L (Lewis 1975: 7).

Deste modo, os membros de P usam a linguagem L quando participam na convenção de sinceridade e confiança em L. Este é o princípio de interpretação a que D. Lewis chama *princípio de sinceridade* (Lewis 1974: 338-39), e afirma que uma frase Ψ é interpretada pela função ϕ tal que, para cada mundo $w \in$ Q, $\phi(w) = 1$, quando cada um dos membros de P, (1) tem um desejo de não dizer Ψ a não ser que Q; (2) acredita que os restantes membros de P têm um desejo semelhante, (3) acredita que Q quando ouve alguém dizer Ψ, (4) acredita que os outros membros de P também acreditam que Q quando ouvem alguém dizer Ψ, (5) acredita que os outros membros de P têm a expectativa de que ele próprio tem as atitudes indicadas em (1)-(4), e, finalmente, (6) acredita que os outros membros de P têm a expectativa de que ele tenha a crença expressa em (5). Sendo isto parte das atitudes de X, fica determinada a interpretação de Ψ para X. Aplicando-se o mesmo método para cada frase usada por X, encontra-se a linguagem por ele usada. (De facto, esta apresentação do princípio da sinceridade é incompleta, porque deixou de lado as frases indexicais. Falarei delas apenas na secção 4.2, depois de ter tratado o tema das crenças *de se*.)

D. Lewis considera que existe ainda um outro princípio de interpretação linguística, o qual designa por *princípio da generatividade*. Este diz-nos que devemos interpretar as frases de X como tendo condições de verdade que sejam finitamente especificáveis e, se possível, de uma maneira uniforme e simples (Lewis 1974: 339).

Ao definir o conceito de linguagem, D. Lewis não restringiu a sua aplicação às linguagens que podem ser usadas por criaturas finitas como os seres humanos, ou cujo uso é capaz de servir os interesses comunicativos dos falantes de uma comunidade. Muitas das linguagens terão um grau de complexidade tão elevado que a sua compreensão é, se não impossível, pelo menos extremamente difícil. E outras serão tão simples que, por esse motivo, são inadequadas para servirem como veículos de informação suficientemente ricos. D. Lewis propõe então que se faça uma distinção no interior das linguagens infinitas (i. e., que contêm infinitas frases) entre aquelas que são, e as que não são, especificáveis por uma gramática:

> «Not just any arbitrary infinite set of verbal expressions will do as the set of sentences of an interesting language. No language adequate to the purposes of its users can be finite; but any language usable by finite human beings must be the next best thing: finitely specifiable. It must have a finite grammar, so that all its sentences, with their interpretations, can be specified by reference to finitely many elementary constituents and finitely many operations for building larger constituents from smaller ones.» (Lewis 1969a: 166-67)

Uma *gramática*, de acordo com D. Lewis, é uma estrutura, finitamente especificável, capaz de gerar as frases de uma linguagem e as suas interpretações a partir de um conjunto finito de expressões elementares e regras para a formação de novas expressões (*ver* secção 1.5.4). Uma gramática G especifica uma linguagem L quando todas as frases construídas de acordo com as regras de G são frases de L e as interpretações dessas frases geradas pelas regras semânticas de G são idênticas àquelas que L lhes atribui. Várias gramáticas podem especificar a mesma linguagem (Lewis 1975: 18). (Encontra-se no anexo 4 uma breve apresentação do género de gramáticas estudado em (Lewis 1970b).)

O princípio de generatividade diz-nos, então, que as condições de verdade das frases da linguagem de X devem ser especificáveis recorrendo a uma gramática. Este é um dos requisitos. O outro diz respeito à uniformidade e simplicidade das gramáticas que têm de ser usadas para tal especificação. Este princípio – e os dois requisitos nele contidos – parecem, à primeira vista, ser desnecessários. Através dos princípios de racionalidade e de caridade, atribui-se conteúdo às atitudes de X e, através destas, são interpretadas as frases por ele usadas recorrendo ao princípio da sinceridade. Como é que uma restrição que exige a conformidade a gramáticas com certas propriedades pode ainda contribuir para a interpretação radical, se esta parece já estar completa?[63]

---

63 Alguém podia ainda apontar que o princípio de generatividade é mesmo pernicioso para uma interpretação, e não apenas desnecessário ou redundante. Este teórico podia pedir-nos para imaginarmos uns entes intelectualmente poderosos – anjos, por exemplo – existentes algures no espaço lógico. Suponhamos que essas criaturas têm linguagem e comunicam entre si, estabelecendo convenções de sinceridade e confiança. A este respeito, seriam como nós. Mas, pela sua condição imaterial, têm o poder de atribuir condições de verdade a um número infinito de frases sem que estas tenham de ser finitamente especificáveis por uma gramática (talvez tenham uma faculdade mental que lhes permite intuir um conjunto infinito de significados a cada momento). Neste caso, os requisitos de D. Lewis levariam a um erro de interpretação: o caminho que propõe é o de proceder a uma simplificação artificial da semântica da linguagem destes entes, em vez de aceitar o facto de estarmos a lidar com criaturas completamente

Em (Lewis 1974), não se encontra qualquer discussão acerca dos benefícios da introdução do princípio da generatividade. No entanto, este parece ser bastante aproximado à solução que, em (Lewis 1992b), D. Lewis considera correta para resolver alguns casos que põem em causa a interpretação linguística através dos padrões associados a uma convenção de sinceridade e confiança. Diz-nos D. Lewis:

> «Consider some very long sentence. Let it be not only long but complicated: clauses within clauses within phrases with clauses …, and abundantly interlaced with cross references to 'the latter', 'the former', 'the aforementioned', 'condition (b*)', and so on *ad nauseam*. Of course you don't expect to hear this sentence uttered. The subjective probability is minute. But what if you hear it? Would you think this was a successful truth-in-L-telling? Not likely! You'd think the speaker was trying to win a bet or set a record, or feigning madness, or raving for real, or doing to annoy, or filibustering, or making an experiment to test the limits of what is humanly possible to say and mean. You wouldn't think he was even trying to be truthful in L. Still less would you think he was trying effectively, armed with skill enough to overcome the complexities of the sentence.» (Lewis 1992b: 108)

O caso com que estamos a lidar é o de uma frase estranha: extremamente longa e complicada, muito dificilmente, ou mesmo impossivelmente, compreensível por nós, dado o insuficiente poder computacional da nossa mente. Provavelmente, esta frase nunca foi dita, e provavelmente nunca o será. Apesar destas características, D. Lewis está a pensar, pelo contexto da discussão, numa frase que intuitivamente classificaríamos como pertencendo à nossa linguagem, já que os seus constituintes são expressões usadas de um modo pouco problemático noutras ocasiões e, de algum modo, esperamos que o significado dessas expressões determine as condições de verdade da frase completa, através de um sistema de regras semânticas estabelecidas por alguma gramática.

No entanto, se o princípio de sinceridade é tudo o que temos em mãos, a frase é excluída da nossa linguagem. Ninguém a usa – nem tem a disposição de a usar em alguma circunstância – como um veículo de informação, esperando que os outros confiem naquilo que está a ser dito. Como nota D. Lewis, o mais provável é que os falantes pensem na

diferentes de nós. (A resposta a esta objeção não pode consistir em dizer que os princípios de interpretação são aplicáveis apenas aos seres humanos e criaturas próximas. Através do problema da interpretação radical pretendemos chegar a uma análise do que é ter atitudes e usar uma linguagem; e essa análise não pode depender de factos contingentes acerca de certos indivíduos.)

elocução desta frase como podendo servir vários propósitos, mas dificilmente aqueles que constituem o uso de uma linguagem. Ou seja, uma frase deste género não está integrada num padrão de sinceridade e confiança necessário para que faça parte da convenção linguística de alguma população.

É aqui que o princípio de generatividade pode ser útil. Por um lado, temos uma restrição que nos diz que a linguagem deve ser finitamente especificável, e que, por isso, todas as suas frases devem poder ser geradas através de um conjunto finito de regras aplicadas a um vocabulário básico finito. Algumas linguagens com frases complicadas e impossíveis de utilizar não são especificáveis dessa maneira e, assim, exclui-se a interpretação que essas linguagens atribuem a frases desse género. Mas, por outro lado, esta restrição ainda não é suficiente. Existem inúmeras gramáticas que podem gerar condições de verdade corretas para todas as frases ditas no mundo atual – no passado, presente e futuro – fornecendo, no entanto, condições de verdade estranhas para o resto das frases. Por exemplo, imaginemos uma gramática que nos diga que qualquer frase em português com mais de um bilião de palavras quer dizer o mesmo que 'Toda a gente é feliz'. Nenhuma destas frases foi ou será dita; por isso, esta gramática especifica as condições de verdade corretas para todas as frases de português usadas pelos seus falantes. Daí a relevância da segunda restrição: a gramática usada para especificar a linguagem deve ser uniforme e simples, o que não acontece com aquela que foi usada no exemplo que acabámos de considerar.

Resumindo, as frases utilizadas, ou pelo menos aquelas acerca das quais os falantes têm certas atitudes, são suficientes para determinar uma classe restrita de gramáticas; de entre estas, as mais simples e uniformes são aquelas que devem ser selecionadas para interpretar o resto da linguagem, por *extrapolação*. É esta a abordagem que D. Lewis apresenta em (Lewis 1992b):

> «First, use somehow determines meaning for the fragment of the language that is actually used. There are rules of syntax and semantics that generate the right sentences with the right meanings within the used fragment. These rules also generate other, longer sentences, with meaning. […] Use determines some meanings, those meanings determine the rules, and the rules determine the rest of the meanings.» (Lewis 1992b: 109)

E, mais à frente, acrescenta:

> «True, there are many grammars [que especificam corretamente as condições de verdade das frases utilizadas]. But they are not on equal terms. Some are 'straight' grammars; for example, any grammar that any linguist would actually propose. Others are 'bent' or 'gruesome', grammars. [...] The notion of extrapolation presupposes the distinction between straight and bent.» (Lewis 1992b: 110)

Aqui, parece haver mais um apelo à naturalidade das propriedades, mas, desta vez, à naturalidade das propriedades das gramáticas – ou, igualmente, às propriedades dos valores semânticos e das regras que compõem as gramáticas. E o propósito é o mesmo que antes: evitar a indeterminação radical, antes do conteúdo mental, agora do significado linguístico.

Capítulo 4 – A intencionalidade *de se* e *de re*

4.1 – As atitudes egocêntricas

As chamadas atitudes intencionais – crenças, desejos, medos, expetativas e intenções, entre outras – têm um certo conteúdo representacional. É comum a ideia de que em todos os casos este conteúdo é apenas *acerca do mundo* – que qualquer crença, por exemplo, representa como verdade que o mundo é desta ou daquela maneira, e nada mais. Em "Attitudes *De Dicto* and *De Se*" (1979a), David Lewis tenta mostrar, através de alguns exemplos, que um sujeito pode ter um conhecimento perfeito acerca do mundo enquanto permanece ignorante relativamente a outras questões. Tendo em conta que ter conhecimento implica ter crença, há por isso mais crenças do que aquelas que dizem alguma coisa sobre o mundo. Um dos exemplos apresentados é este:

> «Consider the case of the two gods. They inhabit a certain possible world, and they know exactly which world it is. […] Still I can imagine them to suffer ignorance: neither one knows which of the two he is. They are not exactly alike. One lives on top of the tallest mountain and throws manna; the other lives on top of the coldest mountain and throws down thunderbolts. Neither one knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws manna or thunderbolts.
> Surely their predicament is possible. (The trouble might perhaps be that they have an equally perfect view of every part of their world, and hence cannot identify the perspectives from which they view it.) […] If the gods came to know which one was which, they would know more than they do.» (Lewis 1979a: 520-21)

Estes deuses são omniscientes relativamente a todos os aspetos do mundo em que habitam. Ao mesmo tempo, no entanto, podem ser completamente ignorantes acerca de que habitantes desse mundo eles são – em particular, cada um deles pode não saber qual dos dois é. A informação que lhes falta não é informação acerca do mundo, mas acerca do lugar que ocupam no mundo. Quando um deles chegar a essa informação, vai ter uma crença correta que é irredutivelmente *egocêntrica*, i. e., acerca de si mesmo e não apenas acerca do mundo. As atitudes unicamente acerca do mundo chamam-se atitudes *de dicto*, e aquelas que dizem respeito ao lugar do sujeito no mundo chamam-se atitudes *de se*.

É assumido que, sendo absolutamente perfeita, a perspetiva de cada um destes deuses relativamente às várias partes da realidade, incluindo os pensamentos e intenções deles próprios, é qualitativamente idêntica, e por isso não há maneira de eles distinguirem qual delas é a sua. A nossa situação cognitiva é, normalmente, bastante diferente desta em que estes deuses se encontram. Temos uma perspetiva parcial do mundo e das coisas que nele existem: temos um acesso imediato apenas àquilo que nos é mais próximo, talvez a partes da nossa vida mental e aos objetos que nos rodeiam. O conhecimento dessa perspetiva é uma marca que permite que nos identifiquemos na rede de coisas que compõem o mundo atual.

O conteúdo da nossa perceção do ambiente circundante é, na maior parte das vezes, egocêntrico. Quando eu vejo um rato a fugir por entre uns arbustos, certamente estou a ver que existem ratos, e que por vezes eles fogem, entre outras coisas acerca do mundo atual. Mas estas são descrições incompletas daquilo que a minha experiência visual representa. O que eu estou a ver é, de facto, que há um rato a fugir por entre uns arbustos, mas também que isso está a acontecer à *minha* frente. Um conteúdo deste tipo é irredutivelmente egocêntrico. Mesmo que eu esteja, por exemplo, nos jardins do Palácio de Cristal, no Porto, não podemos dizer que a minha experiência representa um rato a fugir nos jardins do Palácio de Cristal. Isso seria acrescentar à experiência informação que ela não carrega. Mas dizer que, pelo contrário, a minha experiência representa um rato a fugir, mas em lado nenhum em concreto, seria também incorreto: a minha experiência representa um rato a fugir num lugar concreto – nomeadamente, naquele em que *eu* me encontro. O mesmo se passa com o apercebimento que temos dos nossos próprios estados mentais. Eu apercebo-me, normalmente, que *tenho* uma dor, e não que alguém com estas ou aquelas características tem uma dor.

A gramática superficial da nossa linguagem comum sugere que os objetos das várias atitudes – aquelas coisas que são aceites como verdadeiras, que são desejadas, que são temidas, que são esperadas, que são tencionadas, e por aí em diante – pertencem a diferentes categorias ontológicas. Em (Lewis 1979a: 513), aparecem estes exemplos:

Atribuição de atitude	**Objeto intencional aparente**
"Eu quero *este* gato."	Um gato particular.

“Eu quero um gato qualquer.”	Um gato incompleto.[64]
“Eu quero o inverno.”	Uma época.
“Eu quero uma tempestade.”	Um fenómeno.
“Eu quero escavar na neve.”	Uma atividade.
“Eu quero sentir fadiga.”	Um estado.
“Eu quero que a humanidade habite pelo menos em cinco planetas.”	Uma circunstância (ou uma proposição).

De modo a tornar mais fácil a tarefa de especificar explicitamente o papel causal característico de cada atitude, é conveniente, no entanto, ter objetos intencionais uniformes. Há um número muito grande – talvez infinito – de atitudes e, por isso mesmo, é impossível descrevermos um a um o papel causal que a psicologia popular atribui a cada uma delas. A especificação tem de ser geral. Cada atitude é caracterizada, em parte, por estar integrada numa rede causal com outras atitudes que com ela estabelecem uma certa relação lógica. Podemos, assim, especificar os vários papeis causais dizendo que uma atitude de um certo tipo – uma crença, por exemplo – estabelece esta ou aquela relação causal com outras atitudes que com ela estabelecem esta ou aquela relação lógica. Descrever as relações lógicas entre as atitudes é obviamente mais difícil se tivermos objetos intencionais de uma multiplicidade de categorias.

Exatamente por ser mais conveniente, D. Lewis defende que devemos tratar como ilusória a aparência criada pela linguagem comum. Há, de acordo com ele, uma uniformidade de objetos intencionais mascarada por uma diversidade de maneiras de falar acerca deles (Lewis 1979a: 513-14).

A ideia mais comum é que as *proposições* servem como objetos de todas as atitudes (muitas vezes chamadas, por esse motivo, atitudes *proposicionais*). Existem várias conceções de proposições, e por isso nem toda a gente que identifica os objetos intencionais com as proposições está a dizer exatamente o mesmo. Ainda assim, normalmente concebe-se uma proposição como alguma coisa que pode ser expressa por várias frases e que é verdadeira ou falsa unicamente em virtude da maneira como o mundo

64 Esta é a expressão utilizada por D. Lewis em (1979a: 513) para falar de um estranho gato que, ao contrário dos gatos normais, não tem a especificidade suficiente para ter uma cor ou um tamanho determinado. A única propriedade deste gato é ser um gato. Este tipo de objetos não faz parte da ontologia de D. Lewis e a referência a eles neste contexto serve apenas para ilustrar a aparente variedade dos objetos das atitudes.

é. A teoria, aceite por D. Lewis (*ver* secção 1.5.5), que concebe as proposições como conjuntos de mundos possíveis, exemplifica esta conceção que estamos a considerar.

Associada à tese de que algumas atitudes são irredutivelmente egocêntricas está a tese, também defendida em (Lewis 1979a), de que as proposições, concebidas como conjuntos de mundos, ou pelo menos como portadores de verdade que não sofrem de indexicalidade, não servem como objetos de todas as atitudes.

Parménides acredita que ele mesmo é grego. Qual é o objeto proposicional desta crença *de se*? Presumivelmente, a proposição singular que atribui a Parménides a propriedade de ser grego. Esta é a proposição verdadeira nos mundos em que Parménides (ou, talvez, uma contraparte de Parménides) existe e é grego. Existem vários problemas com esta proposta. Em primeiro lugar, vamos ter de distinguir crenças idênticas. Imaginemos um egípcio louco que também acredita que é grego. Tendo em conta que a crença de Parménides é verdadeira e a do louco é falsa, parece que não têm ambas o mesmo objeto. E é exatamente isso que implica a proposta que estamos a considerar. O objeto proposicional da crença *de se* do louco é a proposição verdadeira nos mundos em que esse louco, e não Parménides, é grego. Agora, imaginemos também que os estados cerebrais do louco ocupam exatamente os mesmos papeis causais que os de Parménides. O propósito de indexar objetos às atitudes, na teoria da mente de D. Lewis, é especificar o papel que eles ocupam na rede causal complexa em que estão integrados. Assim sendo, a diferentes objetos devem corresponder diferentes papeis causais. É errado, por isso, dizer que a crença de Parménides e a do louco têm objetos diferentes. Este exemplo mostra-nos que nenhuma proposição – e não apenas as proposições singulares que estamos a avaliar – pode ser o objeto desta crença *de se*. Estamos a assumir que Parménides e o louco fazem ambos parte do mundo atual, e que a mesma crença é verdadeira num deles e falsa no outro. Mas o valor de verdade de uma proposição não varia entre partes do mesmo mundo.

Em segundo lugar, esta proposta leva a que crenças possíveis se transformem em crenças impossíveis. Alteremos ligeiramente o exemplo e imaginemos que Parménides acredita que é Sócrates. Ele tem então a crença na proposição verdadeira nos mundos em que Parménides é idêntico a Sócrates. Essa proposição, no entanto, é vazia, porque não

há qualquer mundo em que ela é verdadeira. Mas este resultado é absurdo. Aquilo em que Parménides acredita é alguma coisa que pode ser verdadeira – e, atualmente, é verdadeira relativamente a Sócrates. Além disso, quando temos uma crença, restringimos as possibilidades que aceitamos como podendo ser atuais. É precisamente isso que Parménides está a fazer, excluindo a possibilidade de ser alguém diferente de Sócrates. Mas, obviamente, com a crença na proposição vazia não se consegue fazer nada disso (Lewis 1979a: 524-26).

Em terceiro lugar, Parménides pode acreditar que é grego sem acreditar que Parménides é grego. Ele pode, por exemplo, ignorar completamente a sua identidade e desconhecer, por isso, que é Parménides. No entanto, no caso de crenças *de se* terem proposições singulares como objeto, não há diferença nenhuma entre essas duas crenças.

Em quarto lugar, esta proposta torna triviais algumas crenças interessantes. Imaginemos agora que, depois de Parménides estar engando relativamente à sua identidade, ele passou a acreditar corretamente que é mesmo Parménides. Então, *ex hypothesi,* ele passou a acreditar na desinteressante proposição de que Parménides é idêntico a Parménides, o que é absurdo (Lewis 1994: 426).

(Talvez as proposições adequadas às crenças *de se* não sejam proposições singulares. Em vez disso, talvez sejam proposições que envolvem uma descrição – essencial ou acidental – que o sujeito tem de si mesmo. Algumas das objeções que acabamos de ver relativamente à proposta anterior aplicam-se igualmente a esta nova. Parménides acredita na proposição verdadeira nos mundos em que a única pessoa que satisfaz uma certa descrição é grega. O louco acredita no mesmo. Mas o louco está errado e Parménides está certo, apesar de ambos estarem no mesmo mundo. Parménides pode acreditar que é um filósofo sem acreditar que uma pessoa descrita de uma determinada maneira é grega. Além disso, e por fim, Parménides pode ter começado a acreditar que ele mesmo é a pessoa que satisfaz uma certa descrição. Isso significa, inaceitavelmente, que Parménides passou a acreditar na proposição desinteressante de que a pessoa que satisfaz uma certa descrição é a pessoa que satisfaz uma certa descrição.)

Em parte, a proposta de D. Lewis é que as atitudes egocêntricas têm *propriedades* como objetos, e não proposições. Em vez de dizermos que com uma crença *de se* um

sujeito aceita a verdade de uma proposição, devemos antes dizer que ele atribui a si mesmo uma propriedade (Lewis 1979a: 521). Esta proposta resolve os problemas que anteriormente vimos que apareciam com o tratamento dos objetos das crenças *de se* como proposicionais. Consideremos de novo os exemplos que estudámos. (1) Parménides acredita corretamente que é grego, e um louco que também habita o mundo atual acredita falsamente exatamente no mesmo. Não há agora qualquer problema com a diferença no valor de verdade da crença de ambos. Uma propriedade é verdadeira de alguns objetos, e falsa de outros. Ambos acreditam no mesmo, mas Parménides tem uma crença correta porque a propriedade de ser grego é verdadeira dele, e o louco tem uma crença incorreta porque a mesma propriedade não é verdadeira dele. (2) No caso de Parménides acreditar que é idêntico a Sócrates, ele não está a ter uma crença impossível. Existem exemplares da propriedade que é o objeto dessa crença – nomeadamente, Sócrates. (3) Obviamente, Parménides pode acreditar que é grego sem acreditar que Parménides é grego. Seja qual for o tratamento correto desta última crença, ela não é, como a primeira, uma atribuição egocêntrica da propriedade de ser grego. (4) Quando Parménides descobre que é Parménides, ele não passa a acreditar que Parménides é idêntico a Parménides. Presumivelmente, ela já sabia isso antes. O que descobre é que ele próprio tem a propriedade de ser idêntico a Parménides.

Contudo, esta é ainda apenas uma parte da proposta de D. Lewis. Ainda não chegamos à uniformidade de objetos intencionais, como ele pretende. Temos, por um lado, as proposições como objetos das atitudes *de dicto*, e as propriedades como objetos das atitudes *de se*. O resto da proposta é que as atitudes *de dicto* podem ser tratadas como *de se*. (Daí haver uma distinção entre as atitudes *irredutivelmente* egocêntricas, e as restantes, que podem ser consideradas egocêntricas, mas não irredutivelmente.) Quando um sujeito acredita que a neve é branca, podemos dizer que acredita que ele mesmo habita um mundo em que a neve é branca. Em vez de dizermos que ele acredita na proposição verdadeira nos mundos em que neve é branca, dizemos que ele atribui a si mesmo a propriedade de ser um habitante de um mundo em que a neve é branca. De facto, a cada proposição verdadeira numa classe de mundos corresponde uma única propriedade que é verdadeira das partes dos mundos que pertencem a essa classe. Havendo essa

correspondência, temos uma garantia de que é adequado descrever crenças proposicionais como crenças egocêntricas (Lewis 1979a: 516).

A propriedade de habitar um mundo com certas características não captura, certamente, qualquer parte importante do carácter qualitativo de uma coisa. Alguns exemplos tornam isto ainda mais flagrante. Não há necessariamente alguma qualidade relevante comum às coisas que habitam um mundo em que existe um animal com dois corações ou em que há um exemplar de uma molécula de metano. Ainda assim, para D. Lewis, este facto não é razão para contestar a adequação da sua abordagem redutiva às crenças *de dicto*. Uma propriedade é concebida por D. Lewis como um *qualquer* conjunto de objetos possíveis (*ver* secção 1.5.5). Essa é uma conceção abundante de propriedades. Devemos reparar ainda que o conjunto de indivíduos que habitam um mundo com estas ou aquelas características não é uma das propriedades menos naturais e disjuntivas que existem. Pensemos, por exemplo, na propriedade muito mais estranha que tem como membros o primeiro computador que existiu e um lagarto meramente possível.

Como vimos antes, de acordo com D. Lewis, a identificação de certos estados cerebrais com atitudes intencionais numa população é, em parte, aquela que permite dizer que a maioria dos membros da população são racionais – e com isto pretende-se dizer que têm tendência a comportar-se da maneira que melhor serve os seus desejos de acordo com as suas crenças, que vão alterando as suas crenças tendo em conta os estímulos que os afetam, que têm desejos básicos razoáveis, entre outras coisas (*ver* capítulo 3). Essa identificação é, por isso mesmo, holista. Não faz sentido falarmos num estado cerebral como idêntico a uma atitude se ele não estiver integrado numa rede envolvendo outros estados que também se consideram idênticos a outras atitudes. É conveniente, assim, tratarmos do conteúdo de sistemas inteiros de atitudes (crenças e desejos, principalmente), em vez de atitudes isoladas, tendo em conta que é apenas com base nesses sistemas que podemos avaliar a racionalidade de um agente.

Foi precisamente o que fiz no capítulo anterior, quando identifiquei o conteúdo de um sistema de crenças e desejos com duas funções, C e V. A função C distribui parcelas do valor máximo de probabilidade pelos vários mundos. O valor que ela atribui a cada mundo representa a probabilidade de esse mundo ser o atual. A função V, por sua vez,

atribui a cada mundo um número positivo ou negativo que expressa, respetivamente, o grau de desejabilidade ou indesejabilidade de esse mundo ser o atual.

Este método, no entanto, está desenhado para codificar o grau de probabilidade e utilidade que o sujeito atribui a cada proposição, e não a cada propriedade, como acabamos de ver que é necessário. Relembrando, a probabilidade atribuída a uma proposição, concebida como um conjunto de mundos, é a soma da probabilidade atribuída a cada mundo. A utilidade (ignorando agora a complicação de termos de considerar as hipóteses de dependência) de uma proposição, por sua vez, resulta da soma dos produtos da utilidade e da probabilidade condicional de cada mundo relativamente à proposição considerada.

A adaptação deste método à necessidade de acomodar o conteúdo *de se* é, em todo o caso, bastante simples. Enquanto que os pontos do espaço de possibilidades eram antes os mundos, agora são as partes dos mundos (Lewis 1979a: 533-35). Anteriormente, uma região de pontos era uma proposição. Agora, com esta modificação, uma região é uma propriedade, concebida como um conjunto de indivíduos possíveis. Como antes se fazia para as proposições, é possível obter a probabilidade e a utilidade atribuída a cada propriedade.

Normalmente temos crenças parciais. Alguém pode acreditar que é corajoso, mas raramente acredita nisso totalmente, deixando algum espaço à possibilidade de estar errado. Por isso, a soma da probabilidade atribuída a cada membro da classe de pessoas corajosas não vai ser o valor máximo de probabilidade. Mas vamos assumir que um certo sujeito tem apenas crenças totais. Ele vai atribuir a certas propriedades, de um certo inventário, o valor máximo ou o valor nulo de probabilidade. A interseção das propriedades a que atribuiu o valor máximo reúne, numa conjunção, todas as propriedades que ele acredita exemplificar. A cada membro dessa classe chamamos uma *alternativa doxástica* desse sujeito. Estes são os indivíduos que, de acordo com tudo em que ele acredita, podem ser ele mesmo. Aquilo em que ele acredita é aquilo que é verdade de todas as suas alternativas. (Quando estamos a lidar com humanos normais, que não têm crenças totais acerca da maior parte dos temas, podemos continuar a falar em alternativas doxásticas, apesar de ser mais complicado dizer ao certo o que conta como uma

alternativa. Talvez tenhamos de dizer que, neste caso, só há alternativas num certo grau. Ou, de outra maneira, que contam como alternativas aqueles indivíduos que fazem parte do complemento da interseção das propriedades que o sujeito rejeita totalmente exemplificar, ou que acredita não exemplificar num grau muito elevado.)

Um problema para este método é que não permite descrever um sistema com crenças inconsistentes ou desejos conflituantes. Estando confuso, alguém pode acreditar fortemente que é extremamente saudável e, ao mesmo tempo, que está extremamente doente. Antes de mais, será mesmo esta a descrição mais adequada do caso? Não deveríamos antes dizer que esta pessoa está incerta acerca do seu estado de saúde, e por isso atribui um valor reduzido de probabilidade às possibilidades de estar doente e de estar saudável, em vez de estar comprometido com duas certezas contraditórias? Pelo menos de acordo com D. Lewis, devemos mesmo distinguir as duas situações:

> «In your state of doublethink, you have no whole-hearted belief about whether you are healthy; you are half-heartedly certain that you are diseased, half-heartedly certain that you are healthy. The two half-hearted certainties are not at all the same thing as partial belief. Your condition is not one of whole-hearted uncertainty about whether you are diseased or healthy, characterised by one unified system of belief under which some of your alternatives are diseased, some are healthy, and your subjective property is divided more or less evenly between the two subclasses. If you had the opportunity to bet on whether or not you were diseased, the different between the two states would be plain. If you are whole-heartedly uncertain, you hedge your bets. If you are half-heartedly certain each way, you plunge one way or the other – but which you go depends on exactly how the question is put to you, and on how you're feeling at the time.» (Lewis 1986b: 31)

Nesta situação, uma coisa é estar incerto sobre a verdade de P, e atribuir uma probabilidade mais ou menos idêntica a P e não-P, outra é acreditar que P e acreditar que não-P. Uma função C não acomoda esta última situação. A pessoa que acredita que está doente e que está saudável atribui um valor maior do que metade do valor máximo de probabilidade à classe de indivíduos saudáveis e, também, à classe indivíduos doentes. Como estas classes não têm membros em comum, isso significa que a soma do valor atribuído à união das duas classes é maior que o valor máximo de probabilidade, o que é impossível. Quando atribui uma probabilidade muito elevada à classe de saudáveis, a função C deixa a classe de doentes com uma probabilidade baixa, distribuindo quase todo

o valor máximo entre os pontos do espaço de possibilidade que se encontram apenas na região dos saudáveis, deixando sobrar um valor reduzido. E acontece o mesmo inversamente. D. Lewis defende que, nestes casos, devemos atribuir ao sujeito dois sistemas diferentes, em vez de apenas um (Lewis 1986b: 31-3).[65] Evitamos assim ter de colocar as inconsistências e conflitos internos de uma pessoa dentro da descrição de um sistema de crenças e desejos. Essas inconsistências e conflitos são descritos apenas pela constatação de que dois sistemas estão perfeitamente adequados a essa pessoa.

4.2 – Interpretação radical das frases indexicais

O princípio de sinceridade (*ver* secção 3.5) afirma que uma frase ϕ usada por uma população K recebe a intensão (função de mundos para valores de verdade) correspondente à proposição P como interpretação se e só se cada membro de K (1) tem um desejo de não dizer ϕ a não ser que P, (2) acredita que os restantes membros de K têm um desejo semelhante, (3) acredita que P quando ouve alguém dizer ϕ, (4) acredita que os outros membros de K também acreditam que P quando ouvem alguém dizer ϕ, (5) acredita que os outros membros de K têm a expectativa de que ele próprio tenhq as atitudes indicadas em (1)-(4) e, finalmente, (6) acredita que os outros membros de K têm a expetativa de que ele tenha a crença expressa em (5).

Assim formulado, no entanto, este princípio aplica-se apenas a frases eternas, que podem ser interpretadas através de funções de mundos para valores de verdade. As frases indexicais, ao contrário das eternas, têm valores de verdade que variam em diferentes contextos dentro do mesmo mundo. Vimos antes que, para D. Lewis, um contexto pode ser identificado com um falante momentâneo (uma parte temporal de um falante comum)

---

65 Normalmente, temos de dizer que se X acredita que é corajoso e acredita que é famoso, então X também acredita na conjunção dessas propriedades. Ele acredita naquilo que é verdade para todas as suas alternativas. Como ele acredita que é corajoso, todas as suas alternativas são corajosas. E como ele acredita que é famoso, todas as suas alternativas são famosas. Por isso, todas as suas alternativas são corajosas e famosas.

Mas no caso de crenças conjuntamente inconsistentes, isso já não acontece. X acredita que é corajoso e X acredita que não é corajoso. Não é verdade que X acredita que é e não é corajoso. O que é verdade é que ele acredita que é corajoso, de acordo com um dos seus sistemas de crenças, e que não é corajoso, de acordo com outro. Em nenhum deles, no entanto, ele é corajoso e não corajoso. É por isso que lhe podemos atribuir crenças inconsistentes sem lhe imputar uma crença contraditória.

e, por isso, a cada frase – indexical ou eterna – está associada uma função de falantes para valores de verdade (*ver* secção 1.5.4).

Estas funções de contextos para valores de verdade não servem como valores semânticos para as frases, mas servem para reformularmos o princípio da sinceridade de maneira a acomodar as frases indexicais, supondo agora que as atitudes são (às vezes irredutivelmente) egocêntricas e têm propriedades como objetos.

A nova formulação é esta. Uma frase usada por uma população K recebe como interpretação a função de falantes momentâneos para valores de verdade que corresponde à propriedade F (que é o conjunto desses falantes a que a função atribui o valor 1) se e só se cada membro de K (1) tem um desejo de não dizer ϕ a não ser que ele próprio seja F, (2) acredita que os restantes membros de K têm um desejo semelhante, (3) para qualquer falante X pertencente a K, acredita que X é F quando ouve X dizer ϕ, (4) acredita que, para qualquer falante X pertencente a K, os outros membros de K também acreditam que X é F quando ouvem X dizer ϕ, (5) acredita que os outros membros de K têm a expetativa de que de que ele próprio tenha as atitudes indicadas em (1)-(4) e, finalmente, (6) acredita que os outros membros de K têm a expetativa de que ele tenha a crença expressa em (5).

4.3 – Conteúdo restrito e conteúdo lato

Na teoria da mente de D. Lewis, atribuir conteúdo a uma atitude intencional funciona como uma maneira de especificar o papel que essa atitude ocupa numa rede causal complexa que envolve estímulos sensoriais, outros estados mentais e comportamento. A ideia é que cada atribuição de conteúdo capture aquilo em que uma atitude com esse conteúdo difere, em termos de relações causais com o que a rodeia, das restantes atitudes do mesmo tipo. Exemplificando, pretende-se que atribuir a proposição verdadeira nos mundos em que a neve é branca a uma crença permita capturar, juntamente com certos princípios da psicologia popular, o que há de diferente no papel causal da crença de que a neve é branca relativamente ao da crença de que está a chover, ou outra qualquer. Por essa razão, qualquer diferença no conteúdo tem de envolver uma diferença no papel causal, ou estas atribuições deixam de ser funcionalmente adequadas. Uma das características do conteúdo, nesta conceção, é que ele tem de ser *restrito* em todos os

casos, e nunca *lato*. Chamamos restrito ao conteúdo que depende apenas daquilo que ocorre no corpo (ou na alma, se adotarmos uma perspetiva dualista) do sujeito – mais concretamente até, no caso dos homens e outros animais, que depende daquilo que ocorre no seu cérebro ou sistema nervoso central.[66] O mais simples era dizer que um conteúdo é restrito quando ter uma atitude com esse conteúdo é uma propriedade intrínseca do sujeito. Neste contexto, em que estamos a tratar da teoria da mente de D. Lewis, no entanto, essa formulação não é adequada. Em primeiro lugar, provavelmente aquilo que de acordo com esta teoria é relevante para um sujeito ter atitudes com este ou aquele conteúdo – nomeadamente, a ocorrência de estados cerebrais que ocupam este ou aquele papel causal – não faz parte da sua natureza intrínseca. Uma qualidade intrínseca não varia entre duplicados atuais ou meramente possíveis. No entanto, é provável que sujeitos perfeitamente duplicados que vivem em mundos com diferentes leis da natureza podem ter estados funcionalmente diferentes. (Esta ideia é provável, não certa. Imaginemos que as leis da natureza são necessárias. Nesse caso, o que acabei de dizer deixa de ser válido. Não queremos, ainda assim, ficar a depender de uma perspetiva sobre as leis da natureza ou sobre o que é intrínseco para defender a adequação de uma definição de conteúdo restrito.) Em segundo lugar, vimos antes que, algumas vezes, quando um sujeito é um membro estranho da população a que pertence, pode ter um estado mental M porque tem um estado físico F que, apesar de nele não ocupar o papel característico de M, ocupa-o, contudo, na maioria dos membros da população (*ver* secção 2.5).

Por esses dois motivos, o conteúdo restrito não é intrínseco. Ainda assim, uma formulação parecida é adequada. Resolvemos o primeiro problema restringindo a nossa atenção a mundos nomologicamente equivalentes – i. e., com as mesmas leis da natureza. Todos os estados internos de duplicados que existem em mundos nomologicamente

66 O conteúdo tem de ser restrito quando se pretende, como D. Lewis, que as atitudes sejam caracterizadas pelo papel causal que ocupam na mediação conjunta (com outros estados mentais) entre estímulos (*input*) e comportamento (*output*). Esta mediação entre o impacto ambiental e a reação do sujeito ocorre dentro dos limites do organismo, obviamente. Mas existem teorias que exigem descrições funcionais das atitudes que envolvem elementos do ambiente externo. Suponhamos que é uma característica essencial de uma certa atitude, de acordo com uma teoria desse género, que ela seja causada em parte por um elemento químico com certas características, ou pela integração numa sociedade com instituições que funcionam de uma determinada maneira, entre outras coisas. Nestes casos, atitudes devem ser caracterizadas por conteúdo que seja lato, de modo a permitir capturar diferenças nas atitudes que estão muito para além dos limites do organismo do sujeito.

equivalentes ocupam o mesmo papel causal. Resolvemos o segundo problema exigindo que o conteúdo seja intrínseco (ainda com a restrição nomológica) à população inteira a que um sujeito pertence. Temos então esta nova formulação: o conteúdo restrito é aquele que não varia entre sujeitos perfeitamente duplicados integrados em populações intrinsecamente idênticas que vivem em mundos nomologicamente equivalentes. O conteúdo lato é, por sua vez, aquele que pode variar mesmo quando estes aspetos permanecem fixos. Em particular, este tipo de conteúdo não depende apenas das características funcionais dos vários sujeitos, mas também do ambiente – tanto natural como cultural – em que estes se encontram.

Apenas o conteúdo restrito é adequado para diferenciar as atitudes, entendidas como estados funcionalmente caracterizados. Certas experiências mentais, como esta que é apresentada por Hilary Putnam em "Meaning and Reference" (1973) e "The Meaning of 'Meaning'" (1975), no entanto, aparentemente mostram a possibilidade de haver atitudes com um conteúdo lato:

> «For the purpose of the following science-fiction examples, we shall suppose that somewhere there is a planet we shall call Twin Earth. Twin Earth is very much like Earth: in fact, people on Twin Earth even speak *English*. In fact, apart from the differences we shall specify in our science-fiction examples, the reader may suppose that Twin Earth is *exactly* like Earth. He may even suppose that he has a *Doppelganger* – an identical copy – on Twin Earth, if he wishes, although my stories will not depend on this.
> Although some of the people on Twin Earth (say, those who call themselves "Americans" and those who call themselves "Canadians" and those who call themselves "Englishmen," etc.) speak English, there are, not surprisingly, a few tiny differences between the dialects of English spoken on Twin Earth and standard English.
> One of the peculiarities of Twin Earth is that the liquid called "water" is not H_2O but a different liquid whose chemical formula is very long and complicated. I shall abbreviate this chemical formula simply as XYZ. I shall suppose that XYZ is indistinguishable from water at normal temperatures and pressures. Also, I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not water, that it rains XYZ on Twin Earth and not water, etc.
> If a space ship from Earth ever visits Twin Earth, then the supposition at first will be that 'water' has the same meaning on Earth and on Twin Earth. This supposition will be corrected when it is discovered that "water" on Twin Earth is XYZ, and the Earthian space ship will report somewhat as follows.
> "On Twin Earth the word 'water' means XYZ."

> Symmetrically, if a space ship from Twin Earth ever visits Earth, then the supposition at first will be that the word 'water' has the same meaning on Twin Earth and on Earth. This supposition will be corrected when it is discovered that "water" on Earth is H_2O, and the Twin Earthian space ship will report:
> "On Earth the word 'water' means H_2O."
> Note that there is no problem about the extension of the term 'water': the word simply has two different meanings (as we say); in the sense in which it is used on Twin Earth, the sense of water$_{TE}$, what *we* call "water" simply isn't water, while in the sense in which it is used on Earth, the sense of water$_{E}$, what the Twin Earthians call "water" simply isn't water. The extension of 'water' in the sense of water$_{E}$ is the set of all wholes consisting of H_2O molecules, or something like that, the extension of water in the sense of water$_{TE}$ is the set of all wholes consisting of XYZ molecules, or something like that.» (Putnam 1973: 700-01)

A intenção de H. Putnam ao trazer este cenário à consideração é mostrar que, por vezes, o conteúdo de um termo não pode depender unicamente daquilo que têm em mente os membros da comunidade que o usa e, ao mesmo tempo, determinar aquilo que o termo denota tanto no mundo atual como nos restantes meramente possíveis. (Fala-se normalmente da intensão e extensão de um termo. A ideia é que a intensão determina, por vezes em conjunto com o carácter de cada mundo, a extensão do termo nesse mundo. Sendo isto verdade, o exemplo de H. Putnam mostra que a intensão não está no cérebro do falante.) A Terra Gémea é um lugar (no mundo atual ou num mundo meramente possível) semelhante à Terra em quase tudo, mesmo nas características dos seus habitantes (podemos até supor que na Terra Gémea há uma cópia exata de cada habitante da Terra). O único aspeto em que difere da Terra é que nos mares e lagos da Terra Gémea não existe água, mas um líquido superficialmente semelhante, e subjetivamente indiscernível para aqueles que não têm um conhecimento científico aprofundado, que tem a composição química XYZ. Ora, inequivocamente, quando nós usamos o termo 'água' estamos a falar de H2O, mesmo antes de descobrirmos a identidade teórica entre a água e H2O. (Não é razoável supor que os medievais referiam um material diferente ao falar em 'água'.) Pelo contrário, os habitantes da Terra Gémea usam um termo homónimo para falar de XYZ. E continuamos a falar de H2O mesmo quando falamos acerca dos mares e lagos da Terra Gémea dizendo – erroneamente – que neles existe água. Neste exemplo, H2O e XYZ são substâncias indistinguíveis para quem não tem conhecimento acerca da composição química de cada uma. Por isso, antes de chegarmos a esse conhecimento, a

conceção que tínhamos da água era exatamente a mesma que os habitantes da Terra Gémea tinham de XYZ antes de conhecerem a composição química dessa substância. Os terrestres associam aquilo a que chamam 'água' ao que existe nas nuvens, nos mares e nos lagos, que é indispensável à vida, entre outras coisas. Exatamente o mesmo acontece com os terrestres gémeos. Assumindo que termos com o mesmo conteúdo têm o mesmo referente em cada mundo, 'água' na Terra e o homónimo na Terra Gémea têm um conteúdo diferente, tendo em conta que têm um referente diferente no mundo em que está localizada a Terra Gémea, possivelmente o mundo atual. Certo é que o que quer que seja que faça com que o conteúdo desses termos varie nas duas situações não está naquilo que os falantes concebem acerca do referente. O que se passa no cérebro dos terrestres e dos terrestres gémeos é exatamente o mesmo – o conteúdo dos termos depende de fatores externos.

Este cenário pode facilmente ser usado para chegarmos a algumas ideias semelhantes acerca do conteúdo mental. Um de nós que não conhece minimamente a composição química da água tem, ainda assim, inúmeras crenças sobre a água. Ele acredita, por exemplo, que o mar é composto de água. Essa pessoa tem uma cópia exata na Terra Gémea. Essa cópia também tem inúmeras crenças sobre uma substância a que chama 'água'. Uma dessas crenças é, por exemplo, que o mar é composto por essa substância. Ambos têm crenças funcionalmente equivalentes, e ambos expressam o conteúdo das suas crenças através de frases declarativas homónimas usadas exatamente da mesma maneira. Em termos subjetivos, eles vivem em mundos idênticos. Ainda assim, quando o terrestre tem crenças sobre água, a cópia dele na Terra Gémea tem crenças sobre uma outra coisa parecida com água. Quando o terrestre acredita que o mar é composto de água, ele acredita que o mar é composto de H2O. Quem aceitar isto, tem ainda de aceitar que, quando acontece uma situação semelhante com a cópia na Terra Gémea, ela acredita que o mar é composto de XYZ. Conclusão: conteúdo restrito não é suficiente para capturar as diferenças entre aquilo em que os habitantes da Terra e os habitantes da Terra Gémea acreditam.

Com reservas, D. Lewis aceita a descrição que acabei de fazer desta situação. As reservas prendem-se com aquilo que se assume relativamente ao significado do termo

‘água’. Dissemos antes que o nosso termo ‘água’ refere H2O tanto quando falamos da Terra como quando falamos da Terra Gémea. D. Lewis aceita que podemos dizer isto, mas explica que não temos de o fazer obrigatoriamente:

> «Like any up-to-date philosopher of 1955, I think that ‘water’ is a cluster concept. Among the conditions in the cluster are: it is liquid, it is colourless, it is odourless, it supports life. But, pace the philosopher of 1955, there is more to the cluster than that. Another condition in the cluster is: it is a natural kind. Another condition is indexical: it is abundant hereabouts. Another is metalinguistic: many call it ‘water’. Another is both metalinguistic and indexical: I have heard of it under the name ‘water’. When we hear that XYZ off on Twin Earth fits many of the conditions in the cluster but not all, we are in a state of semantic indecision about whether it deserves the name ‘water’. […] When in a state of semantic indecision, we are often glad to go either way, and accommodate our own usage temporarily to the whims of our conversational partners (Lewis 1979b). So if some philosopher, call him Schmutnam, invites us to join him in saying that the water on Twin Earth differs in chemical composition from the water here, we will happily follow his lead. And if another philosopher, Putnam (1975), invites us to say that the stuff on Twin Earth is not water – and hence Twoscar does not believe that water falls from clouds – we will just as happily follow his lead. We should have followed Putnam’s lead only for the duration of that conversation, then lapsed back into our accommodating state of indecision. But, sad to say, we thought that instead of playing along with a whim, we were settling a question once and for all. And so we came away misled.» (Lewis 1994: 424)

Nesta passagem, D. Lewis está a apresentar a variedade de descritivismo acerca da referência que considera mais adequada. Contrariamente aos proponentes da teoria descritivista clássica, D. Lewis considera que o feixe de descrições que compõem o conceito associado a um termo não deve incluir apenas algumas características do referente. Consideremos o exemplo de ‘Aristóteles’. Um feixe que inclua apenas descrições como ‘o professor de Alexandre Magno’ e ‘o mais famoso aluno de Platão’ não serve para determinar adequadamente o referente do termo nos vários mundos. Temos ainda de acrescentar condições egocêntricas que capturem a maneira como o falante está relacionado com o referente através de uma cadeia causal. A relação pode ser percetual, como em ‘o filósofo que estou a ver neste momento’ (uma relação que não podemos mais estabelecer com Aristóteles). Ou pode ser metalinguística, como em ‘a pessoa de quem ouvi falar através do nome “Aristóteles”’. (Pode também ser usada a descrição ‘a pessoa a que chamam “Aristóteles”’, bastante parecida com esta, mas que já não inclui o

elemento egocêntrico.) Estas modificações à teoria clássica tentam responder às objeções contra o descritivismo levantadas em (Kripke 1980). Uma dessa objeções é que descrições (ou feixes de descrições) não têm o mesmo comportamento modal que os nomes aos quais supostamente fornecem o significado. Imaginemos que a determinação do referente de 'Aristóteles' é conseguida através da descrição 'o professor de Alexandre Magno'. (Acabamos de ver que muito provavelmente isso não é verdade, mas as modificações necessárias são complicações que não alteram aquilo que vou dizer.) O que havemos de dizer então acerca do referente de 'Aristóteles' num mundo em que o professor de Alexandre Magno é, por exemplo, Parménides? O referente de 'o professor de Alexandre Magno' é Parménides, por isso o de 'Aristóteles' também deve ser. O filósofo a que chamamos no mundo atual 'Aristóteles' até pode existir nesse mundo, mas aí não é nomeado por 'Aristóteles'. Esta ideia é obviamente errada. Basta repararmos que aceitá-la tem como consequência que a frase 'Aristóteles é o professor de Alexandre Magno' é necessariamente verdadeira, quando supomos que expressa uma questão de facto contingente. Dizemos então, pelo contrário, que 'Aristóteles' é um designador rígido, e refere Aristóteles em todos os mundos em que ele existe. É, assim, preciso que possamos rigidificar o feixe de descrições associadas a um termo. 'Aristóteles' não é sinónimo de 'o professor de Alexandre Magno e ...', mas sim 'aquilo que *no mundo atual* é o professor de Alexandre Magno e ...' (Lewis 1984: 223). (A rigidificação pode envolver ainda lugares mais restritos que o mundo atual. Podemos dizer que o referente é aquilo que satisfaz uma certa descrição aqui nas nossas redondezas. Um habitante da Terra vai conceber o referente de 'água' rigidamente como o líquido com estas ou aquelas características que existe *à volta dele*, sendo para ele irrelevante a existência de um líquido superficialmente semelhante noutra região do mundo atual, como aconteceria se a Terra Gémea fosse um planeta atual.)[67]

[67] Podemos capturar a diferença entre o conceito associado a um termo e o comportamento que ele apresenta em contextos modais caracterizando o seu significado através do método, que D. Lewis defende em (Lewis 1980a), de inserir nos valores semânticos um contexto e um indíce (*ver* também a secção 1.5.4). Variando o contexto, podemos ter referentes variáveis. 'Aristóteles' refere Aristóteles num contexto atual, mas Parménides num mundo meramente possível em que Parménides é o professor de Alexandre Magno. Essa variação, no entanto, não afeta as frases com operadores modais, porque o que é relevante para a avaliação da verdade dessas frases é aquilo que o termo refere se mudarmos a coordenada modal do índice (o mundo possível), mantendo fixo o contexto. Mantendo um contexto atual, 'Aristóteles' refere

Esta teoria tem consequências para a maneira como D. Lewis pensa no caso da Terra Gémea. Aquilo a que os habitantes da Terra Gémea chamam 'água' satisfaz algumas das características que associamos à água. No entanto, não é aquilo que satisfaz as nossas descrições egocêntricas e metalinguísticas. Não é aquilo com que nós mantemos uma certa relação percetual e chamamos 'água', por exemplo. Temos então um exemplo de indecisão semântica. Atribuindo uma maior importância a uma das partes do conceito de 'água' vamos dizer que os mares da Terra Gémea não são compostos por água. Atribuindo uma maior importância a outra parte desse conceito, já vamos dizer isso. Optando por esta última alternativa, vamos também dizer que os habitantes da Terra Gémea têm crenças acerca da água quando estão num estado funcionalmente equivalente àquele em que estamos nós quando temos crenças sobre a água. Ambas as alternativas são aceitáveis. Desse modo, é aceitável falar como era proposto por H. Putnam. Numa certa desambiguação, então, é verdade que um sujeito X na Terra acredita que o mar é composto de água e, ao mesmo tempo, é também verdade que uma cópia dele na Terra Gémea, com estados internos que ocupam exatamente o mesmo papel causal que os correspondentes estados em X, não acredita nisso. A diferença entre os dois reside naquilo que os rodeia. O conteúdo que acabamos de atribuir a X e à sua cópia é lato, não restrito.

Em "Individualism and the Mental" (1979: 77-9), Tyler Burge utiliza a seguinte experiência de pensamento de maneira a argumentar contra uma conceção "individualista" da mente. Uma pessoa X tem várias atitudes que podem ser corretamente atribuídas através de frases que contêm o termo 'artrite'. Por exemplo, X acredita que sofre de artrite, que a artrite que tem nos pulsos e nos dedos é mais dolorosa que a artrite que tem nos tornozelos, que é melhor sofrer de artrite que de cancro do fígado, que o endurecimento das articulações é um sintoma de artrite, que certas dores são

rigidamente Aristóteles em qualquer situação contrafactual. É por isso que a frase 'Aristóteles é o professor de Alexandre Magno' não é necessariamente verdadeira, apesar de ser verdadeira, na hipótese de definirmos assim 'Aristóteles', em todos os contextos em que é dita. Esta ideia está na base das semânticas bidimensionais, que atribuem a um termo duas funções de mundos para referentes. Uma delas (a intensão primária) é a função de mundos para o referente que o termo teria caso a sua referência fosse fixada nesse mundo – i. e., caso o termo estivesse a ser usado nesse mundo. A outra (a intensão secundária) é a função de mundos para o referente que o termo tem nesse mundo, estando fixado o seu uso no mundo atual (Chalmers 1996: 2.6). A intensão primária permite descrever uma componente *a priori* do significado, que é perdida com a intensão secundária.

características da artrite, que há vários tipos de artrite, entre outras coisas. Além disso, acredita – falsamente – que sofre de artrite na coxa. A certa altura, X conta a um médico essa suspeita, que o corrige imediatamente, explicando-lhe que é impossível sofrer uma atrite na coxa porque o termo 'artrite' aplica-se apenas a uma inflamação nas articulações. O erro de X é, em parte, um erro de compreensão de 'artrite'.

Imaginemos agora um outro mundo possível nomologicamente equivalente ao atual em que existe uma contraparte de X chamada X*, com exatamente a mesma natureza intrínseca que X até ao momento em que ele fala com o médico sobre a suposta artrite na coxa. X e X* têm, até esse momento, a mesma história fisiológica, as mesmas características físicas e funcionais, as mesmas disposições verbais e não-verbais, foram afetados pelos mesmos estímulos e envolveram-se nos mesmos comportamentos, entre outras coisas. Em particular, X e X* usam o termo 'artrite' da mesma maneira. Exatamente as mesmas circunstâncias levam ambos – atual ou contrafactualmente – a enunciarem uma frase com esse termo, e pelo menos enquanto a informação é alguma coisa que depende daquilo que se passa num cérebro, ambos associam a mesma informação a 'artrite'. Ainda assim, a comunidade a que X* pertence usa o termo 'artrite' para falar tanto de uma inflamação nas articulações, como de outras doenças, incluindo uma inflamação nas coxas, e não usa mais nenhum termo para falar especificamente de artrite. Podemos dizer, por isso, que X*, contrariamente a X, não sofre de qualquer incompreensão acerca do significado do termo 'artrite'.

Assumimos antes que X tinha inúmeras crenças sobre a artrite, e isso não parece ter ficado controverso a partir do momento em que descobrimos que ele sofre de uma compreensão errada do termo 'artrite'. Não é certo, no entanto, que possamos dizer o mesmo de X*. É verdade que ambos adquirem crenças acerca daquilo a que 'artrite' se aplica por terem esse termo no seu vocabulário, e é verdade também que ambos usam esse termo exatamente da mesma maneira, mas a diferença no conteúdo que a comunidade de cada um deles atribui a 'artrite' parece afetar o conteúdo das suas crenças. Uma das formas de conseguir pensar sobre alguma coisa, presumivelmente, passa por adquirir uma palavra que a denota. Para isso, ninguém tem de ter capturado todas as nuances do uso correto dessa palavra – a comunidade em geral, de certa maneira, pode

cumprir essa tarefa por cada um dos seus membros. É por isso que é incontroverso que X tem crenças sobre artrite, utilizando para isso um termo que a comunidade aplica a uma inflamação nas articulações. E é também por isso que é incontroverso que X*, quando tem estados mentais semelhantes aos que em X se referem a artrite, está a pensar numa coisa ligeiramente diferente. A comunidade dele – e, no caso, ele próprio – fazem uso de 'artrite' para falar de artrite e de mais algumas inflamações, incluindo na coxa. É sobre isso, chamemos-lhe artrite*, que X* está apto a pensar.

É isto que T. Burge conclui deste exemplo:

> «The upshot of these reflections is that the patient's mental contents differ while his entire physical and non-intentional mental histories, considered in isolation from their social context, remain the same. (We could have supposed that he dropped dead at the time he first expressed his fear to the doctor.) The differences seem to stem from differences "outside" the patient considered as an isolated physical organism, causal mechanism, or seat of consciousness. The difference in his mental contents is attributable to differences in his social environment. In sum, the patient's internal qualitative experiences, his physiological states and events, his behaviourally described stimuli and responses, his dispositions to behave, and whatever sequences of states (non-intentionally described) mediated his input and output – all these remain constant, while his attitude contents differ, even in the extensions of counterpart notions. As we observed at the outset, such differences are ordinarily taken to spell differences in mental states and events.» (Burge 1979: 79)

Aproximadamente, estamos perante um caso semelhante ao da Terra Gémea. Em ambos, temos um sujeito X no mundo atual com uma natureza intrínseca N e com crenças acerca de uma coisa Y, e temos um sujeito X* num mundo meramente possível nomologicamente equivalente ao atual (ou pelo menos numa zona remota do mundo atual) que exemplifica N, mas que, por um motivo ou outro, não tem crenças acerca de Y, mas acerca de uma outra coisa que aparece a X* de uma maneira semelhante àquela como Y aparece a X. A explicação da diferença, seja ela qual for, tem de estar fora daquilo que é abrangido pela natureza intrínseca N integrada em mundos com certas leis da natureza.

Há, no entanto, uma diferença a ter em conta. No exemplo da Terra Gémea, aquilo que explicava a diferença no conteúdo era apenas uma diferença nos componentes do ambiente natural que rodeavam cada um dos sujeitos e as suas populações, e nada que

acontecia com estes – até ao ponto de o exemplo funcionar mesmo assumindo que todos os habitantes da Terra têm um duplicado perfeito na Terra Gémea. O mesmo não acontece agora com o caso da artrite. A diferença no conteúdo resulta de uma diferença na prática linguística dos membros da comunidade de cada sujeito. Uma diferença nesse aspeto de várias populações que habitam mundos em quase tudo semelhantes implica certamente uma diferença nas qualidades intrínsecas de alguns dos respetivos membros. Duas consequências se seguem daqui. Em primeiro lugar, variações no conteúdo podem resultar tanto de fatores naturais como de diferenças no ambiente social e cultural de um sujeito, mesmo quando o conteúdo diz respeito a aspetos da realidade natural (não-social e não-cultural), como é o caso das ocorrências de uma doença como a artrite.

Em segundo lugar, a definição de conteúdo restrito utilizada anteriormente não é adequada. Foi útil para tratar do caso da Terra Gémea, e seria útil para tratar também outros casos envolvendo diferenças a nível de fatores naturais, mas não casos deste género que estamos agora a considerar. O conteúdo restrito foi definido como o conteúdo que não varia entre sujeitos perfeitamente duplicados que estão integrados em populações também perfeitamente duplicadas e que habitam mundos com as mesmas leis da natureza. Acontece que, por esta definição, o conteúdo de X acerca da artrite é restrito. Mantendo intacta a natureza intrínseca da população de X, é impossível chegar a uma situação em que uma contraparte de X não pensa em artrite porque a população dela não fala de artrite quando usa o termo 'artrite'. Por isso, esta definição não captura, como me parece que deve, o carácter individualista que o conteúdo restrito tem, e que T. Burge pretende denunciar como inadequado para uma teoria geral da mente.[68] Aqui está uma nova proposta: o conteúdo restrito é aquele que não varia entre sujeitos perfeitamente

[68] É verdade que a teoria da mente de D. Lewis é, de certa forma, comunitarista, como mostra o caso da dor louca (*ver* secção 2.5). Não é de admirar, por isso, que o conteúdo, de acordo com D. Lewis, não seja determinado totalmente de uma maneira individualista. No entanto, essa teoria apresenta uma conceção de mente muito mais individualista do que aquela que T. Burge pretende defender em (Burge 1979), e é precisamente essa diferença que a definição de conteúdo restrito que antes ofereci não captura. Para D. Lewis, a possibilidade de o conteúdo depender da comunidade é excecional. A inadequação do papel causal de um estado físico ao estado mental a que é idêntico tem de ser uma situação atípica, sob pena de esse estado físico não ser identificável mais com esse estado mental. A mesma atitude, na maior parte dos outros casos, depende apenas do papel que desempenha na rede causal dos estados do sujeito. Em T. Burge, no entanto, a ideia é que há que ter em conta a interação de um indivíduo com a comunidade para ter uma teoria correta do conteúdo, e que a dependência da comunidade não pode ser encarada – e, assim, desvalorizada – como uma anomalia.

duplicados que estejam integrados em populações onde nenhum membro é atípico no que diz respeito à ocupação de papeis causais por parte dos seus estados (cerebrais ou de outro tipo). Com esta clausula, é possível evitarmos a situação em que um sujeito tem uma atitude com um certo conteúdo por causa daquilo que ocorre com os outros membros da população, a qual motivou a definição anterior.

Com o exemplo da Terra Gémea (e o da artrite), percebemos que estes três enunciados podem ser ao mesmo tempo verdadeiros, pelo menos numa certa desambiguação:

(1) X acredita que as nuvens são compostas por H2O.

((1*) X acredita que sofre de artrite.)

(2) Y não acredita que as nuvens são compostas por H2O.

((2*) Y não acredita que sofre de artrite.)

(3) X e Y têm um cérebro (ou sistema nervoso central) funcionalmente equivalente.

(Estou a ignorar, propositadamente, possíveis casos em que as comunidades de X e Y têm membros atípicos, de modo a tornar a apresentação mais fácil. Nada do que à frente é dito teria de ser significativamente diferente sem esta restrição.)

Aparentemente, a possibilidade da verdade em conjunto de (1), (2) e (3) (e, igualmente, de (1*), (2*) e (3)) é incompatível com esta ideia central na teoria da mente de D. Lewis:

(4) É analiticamente verdade que se X tem um estado físico F que ocupa o papel causal característico do estado mental M, então M = F e, por isso, X tem M.

Essa aparência aproxima-se da realidade, mas não totalmente. A incompatibilidade surge de uma maneira óbvia apenas quando (1), (2) e (3), assim como (1*) e (2*), são analisadas deste modo:

(5) ‘X acredita que F(A)’ (quando ‘A’ é um nome próprio e ‘F’ é um predicado) serve para atribuir a X uma crença que tem como objeto a proposição expressa no mundo atual pela frase ‘F(A)’.

(6) Essa proposição é verdadeira nos mundos em que o objeto denotado no mundo atual por ‘A’ (ou uma contraparte desse objeto) tem a propriedade expressa no mundo atual por ‘F’.

Esta análise parte de duas ideias que vou assumir como corretas. Uma delas é que uma proposição é individuada através dos mundos possíveis em que é verdadeira, como acontece, por exemplo, se for identificada com o conjunto desses mundos. A outra tem a ver com a maneira como uma frase exprime uma proposição. É dito em (6) que uma frase que atribui uma propriedade a um objeto (e o mesmo pode ser dito, *mutatis mutandis*, de uma frase que atribui uma relação a uma sequência deles) exprime a proposição que é verdadeira nos mundos em que *esse* objeto (ou, de novo, uma contraparte *desse* objeto) tem essa propriedade. Relembrando um exemplo que encontramos em (Frege 1892), isto leva a que um par de frases como ‘A Estrela de Manhã é Vénus’ e ‘A Estrela da Tarde é Vénus’ estejam associadas à mesma proposição, tendo em conta que ‘Estrela da Manhã’ e ‘Estrela da Tarde’ têm o mesmo referente – nomeadamente, o planeta Vénus. A ideia de Gottlob Frege e, em geral, dos proponentes das teorias descritivistas da referência é que o conteúdo destas frases não é o mesmo, porque, apesar de serem nomes correferentes, ‘Estrela da Manhã’ e ‘Estrela da Tarde’ estão associados a uma diferente conceção (ou descrição) do objeto que denotam, e assim cada um contribui de uma maneira específica para aquilo que exprimem as frases que se constroem a partir deles. Uma forma de tornar mais precisa esta ideia passa por identificar o conteúdo de um nome (entendido esse conteúdo como o contributo composicional do nome) com uma função que liga cada mundo ao objeto, quando existe, que o nome refere nesse mundo. Este método permite diferenciar o conteúdo de nomes que têm o mesmo referente no mundo atual, mas não em todos os restantes mundos. Quanto a uma frase, identifica-se o seu conteúdo com uma função de mundos para valores de verdade. É natural dizer, a partir disto, que o conjunto de mundos em que a frase é verdadeira é idêntico à proposição

expressa pela frase. Ainda assim, considerações como aquelas que encontramos em (Kripke 1980) mostraram que, com uma enorme probabilidade, este método é inadequado, pelo menos se um dos objetivos era que conseguíssemos através dele capturar o que há de diferente naquilo que é dito pelas frases 'A Estrela da Manhã é Vénus' e 'A Estrela da Tarde é Vénus'. 'Estrela da Manhã' e 'Estrela da Tarde', contrariamente ao que era exigido para podermos afirmar que diferem em conteúdo, têm o mesmo referente em todos os mundos em que têm um referente. (Não assumir isto teria consequências desastrosas para o tratamento de frases com operadores modais.) Uma das coisas que podemos dizer, perante isto, é que podemos distinguir entre o referente que um nome tem nos vários mundos estando a ser usado no mundo atual, e o referente que ele teria caso fosse usado noutros mundos. O que S. Kripke mostrou é que, fixado num certo mundo, o referente tem de ser o mesmo em todos os mundos considerados a partir daí, mas não que o referente tem de ser o mesmo caso o mundo em que o nome está a ser usado fosse outro. Estes dois tipos de padrões de avaliação podem capturar diferentes aspetos do conteúdo de um termo, e conciliar algumas das ideias de G. Frege com as descobertas de S. Kripke. Consequência desta preferência é que uma frase como 'F(A)' tem agora duas proposições como candidatas a serem a proposição expressa por ela. Uma delas é a proposição verdadeira nos mundos em que o objeto denotado por 'A' no mundo atual tem a propriedade F, a outra é a proposição verdadeira nos mundos em que o objeto que seria denotado por 'A' se usado nesse mundo tem a propriedade F. A análise que estamos a considerar tem por base a primeira opção, e por isso nada do que será dito se aplica diretamente a uma análise que preferisse a última.

Esta análise parte ainda de uma outra ideia, relativamente comum, de que podemos individuar as crenças de um sujeito através das frases verdadeiras que lhe atribuem uma crença com um certo conteúdo, mais ou menos como é descrito por T. Burge em "Individualism and the Mental":

> «Thoughts, beliefs, intentions, and so forth are typically specified in terms of subordinate sentential clauses, that-clauses, which may be judged as true or false. Pains, feels, tickles, and so forth have no special semantical relation to sentences or to truth or falsity. […]

> In an ordinary sense, the noun phrases that embed sentential expressions in mentalistic idioms provide the *content* of the mental state or event. We shall call that-clauses and their grammatical variants "*content-clauses*." Thus the expression 'that sofas are more comfortable than pews' provides the content of Alfred's belief that sofas are more comfortable than pews. […]
> The crucial point in the preceding discussion is the assumption that obliquely occurring expressions in content clauses are a primary means of identifying a person's intentional mental states or events.» (Burge 1979: 74)

Em "What Puzzling Pierre Does Not Believe" (1981b), D. Lewis argumenta que esta análise é refutada pelo hipotético caso de Pierre, apresentado por S. Kripke em "A Puzzle about Belief" (1979).[69] Este é o caso:

> «Suppose Pierre is a normal French speaker who lives in France and speaks not a word of English or of any other language except French. Of course he has heard of that famous distant city, London (which he of course calls '*Londres*') though he himself has never left France. On the basis of what he has heard of London, he is inclined to think that it is pretty. So he says, in French, "*Londres est jolie*."
> On the basis of his sincere French utterance, we will conclude:
>
> (7) Pierre believes that London is pretty.
>
> I am supposing that Pierre satisfies all criteria for being a normal French speaker, in particular, that he satisfies whatever criteria we usually use to judge that a Frenchman (correctly) uses '*est jolie*' to attribute pulchritude and uses '*Londres*' – standardly – as a name of London.
> Later, Pierre, through fortunate or unfortunate vicissitudes, moves to England, in fact to London itself, though to an unattractive part of the city with fairly uneducated inhabitants. He, like most of his neighbors, rarely ever leaves this part of the city. None of his neighbors know any French, so he must learn English by 'direct method,' without using any translation of English into French: by talking and mixing with the people he eventually begins to pick up English. In particular, everyone speaks of the city, 'London,' where they all live. Let us suppose for the moment – though we will see below that this is not crucial – that the local population are so uneducated that they know few of the facts that Pierre heard about London in France. Pierre learns from them everything they know about London, but there is little overlap with what he heard before. He learns, of course – speaking English – to call the city he lives in 'London.' Pierre's surroundings are, as I said, unattractive, and he is unimpressed with most of the rest of what he happens to see. So he is inclined to assent to the English sentence:
>
> (8) London is not pretty.

69 Falando mais rigorosamente, D. Lewis argumenta que esta análise falha se a ela lhe juntarmos a ideia nada controversa para quem aceita o tratamento das proposições como conjuntos de mundos possíveis, de que uma classe de crenças é inconsistente se não existir nenhum mundo possível em que os seus objetos proposicionais são todos verdadeiros. Como não é razoável negar isto, o problema com a análise, a haver, vai ter de residir obviamente em (5) ou (6).

> He has no inclination to assent to:
>
> (9) London is pretty.
>
> Of course he does not for a moment withdraw his assent from the French sentence, "*Londres est jolie*"; he merely takes it for granted that the ugly city in which he is now stuck is distinct from the enchanting city he heard about in France. But he has no inclination to change his mind for a moment about the city he still calls '*Londres*.'
>
> This, then, is the puzzle. If we consider Pierre's past background as a French speaker, his entire linguistic behavior, on the same basis as we would draw such a conclusion about many of his countrymen, supports the conclusion ((7) above) that he believes that London is pretty, On the other hand, after Pierre lived in London for some time, he did not differ from his neighbors – his French background aside – either in his knowledge of English or in his command of the relevant facts of local geography. His English vocabulary differs little from that of his neighbors. He, like them, rarely ventures from the dismal quarter of the city in which they all live. He, like them, knows that the city he lives in is called 'London' and knows a few other facts. Now Pierre's neighbors would surely be said to use 'London' as a name for London and to speak English. Since, as an English speaker, he does not differ at all from them, we should say the same of him. But then, on the basis of his sincere assent to (8), we should conclude:
>
> (10) Pierre believes that London is not pretty.» (Kripke 1979: 254-56)[70]

Supomos que Pierre não tem crenças conjuntamente inconsistentes. Ele acredita, é verdade, que Londres é bela, e também que Londres não é bela, mas não tem meios disponíveis para detetar, e corrigir, essa aparente contradição. A não ser que obtenha mais informação sobre o mundo – em particular, informação acerca do referente de 'London' e 'Londres' – Pierre vai permanecer incapaz de perceber o (suposto) erro lógico que (supostamente) está a cometer. E nova informação serve para perceber o que é verdadeiro e o que é falso, mas não para detetar uma contradição previamente existente. Como afirma S. Kripke, depois de discutir algumas alternativas de interpretação deste caso:

> «[...] surely anyone, leading logician or no, is in principle in a position to notice and correct contradictory beliefs if he has them. Precisely for this reason, we regard individuals who contradict themselves as subject to greater censure than those who merely have false beliefs. But it is clear that Pierre, as long as he is unaware that the cities he calls 'London' and *'Londres*' are one and the same, is in no position to see, by logic alone, that at least one of his beliefs must be false.

70 Acomodei a numeração dos enunciados deste excerto ((7), (8) e (9)) à numeração geral desta secção.

> He lacks information, not logical acumen. He cannot be convicted of inconsistency: to do so is incorrect.» (Kripke 1979: 257)

Não há, no entanto, qualquer mundo possível em que Londres (ou uma contraparte de Londres) é e não é bela. E, de acordo com a análise proposta em (5) e (6), (7) atribui a Pierre a crença que tem como objeto a proposição verdadeira nos mundos em que Londres é bela, e (10) a crença na verdade da proposição verdadeira nos mundos em que Londres não é bela. A análise tem de estar errada nalgum ponto. D. Lewis vai rejeitar o que é dito em (5). Algumas frases, contrariamente ao que parece, são apenas em parte acerca da vida psicológica de um sujeito. Dizer que X acredita na verdade da proposição P não é, por isso, o mesmo que dizer que X tem – dentro do cérebro – um estado físico que pode ser identificado com a crença na verdade de P. É verdade que Pierre acredita que Londres é bela e que Londres não é bela. De acordo com D. Lewis, no entanto, isso não significa que é verdade que Pierre tenha a crença de que Londres é bela e tenha outra crença de que Londres não é bela. Ele acredita que Londres é bela e que Londres não é bela em parte pelas crenças que tem e, em parte, por alguma coisa que se passa ao seu redor. Ele não consegue detetar por meios lógicos a contradição inerente àquilo em que acredita, porque a responsabilidade dessa contradição é daquilo que acontece fora da mente dele – e é por essa mesma razão que o que ele precisa para sair da inconsistência é de informação, e não de técnicas lógicas.

Há assim uma separação entre as atitudes e as frases da linguagem comum acerca delas: «there are various ways for a system of belief to make a belief sentence true. I cannot propose a uniform formula to cover all cases.» (Lewis 1986b: 32) Uma coisa é o sistema, caracterizado funcionalmente, de crenças e de outras atitudes implementado no cérebro de um sujeito. Outra coisa diferente – e, podemos acrescentar, sem uma relação uniforme e constante – é a verdade das atribuições de atitudes a esse sujeito que se fazem através de frases que envolvem operadores como ‘X acredita que’, ‘X quer que’, ‘X teme que’, ‘X espera que’ e ‘X tenciona fazer com que ...’, entre outros. Em certos casos, a conexão entre as duas coisas é bastante simples. X tem um estado interno que ocupa o papel causal que faz com que seja idêntico à crença na verdade da proposição de que a felicidade é mais importante que a virtude. É por esse motivo que é verdadeira a frase ‘X

acredita que a felicidade é mais importante que a virtude'. Mas algumas das atribuições vão mais longe, e não se limitam a falar daquilo que ocorre na vida mental de um sujeito.

Esta ideia permite a D. Lewis distinguir entre um sentido básico de crença e de conteúdo, e um sentido derivado dessas coisas. Neste sentido derivado, dizemos que um sujeito tem uma crença com um certo conteúdo quando é verdadeira dele a frase que atribui a crença com esse conteúdo. É apenas de uma maneira derivada que um sujeito tem crenças com conteúdo lato. Os habitantes da Terra e os habitantes da Terra Gémea acreditam, no sentido básico, exatamente no mesmo, porque têm estados internos que ocupam os mesmos papeis causais. Mas, derivadamente, acreditam em coisas diferentes.

Assim, o conteúdo restrito e o conteúdo lato operam em tarefas diferentes. O conteúdo restrito, como vimos anteriormente, serve para caracterizar unicamente as diferenças entre papeis causais e, por isso mesmo, é o conteúdo que captura aquilo que é essencial a cada atitude. E o conteúdo lato? «[w]ide content does serve a purpose. It enters into the analysis of sentences that are about belief, or at least partly about belief; or at least it does so under some permissible disambiguations of these sentences.» (Lewis 1994: 424) Resta ainda ver como quais são os contornos gerais desta análise.

4.4 – Atitudes *de re* e o conteúdo lato

Chamamos *de re* a crenças que servem para atribuir uma propriedade a um certo objeto (ou uma relação a uma sequência deles). Aparentemente, crenças *de dicto* subsumem as *de re*. Atribuir a propriedade F ao objeto A é basicamente ter uma crença *de dicto* na proposição verdadeira nos mundos em que A (ou uma contraparte de A) existe e tem a propriedade F. (E ter uma crença *de dicto* na verdade de uma proposição é ter um estado cerebral que ocupa o papel causal característico de uma crença com esse conteúdo.) Esta ideia é simples, mas errada. Inequivocamente, Pierre atribui beleza a Londres (e também lhe atribui fealdade). Ainda assim:

> «Pierre does not have as an object of his belief the proposition (actually) expressed by 'London is pretty'. For there is a possible world which fits Pierre's beliefs perfectly – it is one of his 'belief worlds' – at which the proposition is false.
> I have in mind a world where the beautiful city Pierre heard about was not London but Bristol. Imagine a world just like ours until fairly recently (except to the extent

> that it must differ to fit Pierre's misconception about earlier history, if any). Then the beautification of Bristol was undertaken, and at the same time it was renamed in honour of Sir Ogdred Londer. The French called this famous city '*Londres*'; they spoke often of its beauty, and all they said was true. In due course Pierre heard of the beauty of Bristol, lately called 'Londer' in England and '*Londres*' in France, and he came to assent sincerely to '*Londres est jolie*'. What happened at this end was just like what happened at the real world.
> While Bristol was beautified, London fell into decay. The better parts were demolished – copies sometimes were built in Bristol, alias 'Londer' – and only the slums remained. London became ugly through and through. Also, nothing of consequence happened there. The French had little occasion to speak of the place under any name, and indeed it never was mentioned in Pierre's presence. It was to this place that the unfortunate Pierre was made to go. Again, what happened at Pierre's end of his encounters with London was just like what happened at the real world.
> This world fits Pierre's beliefs perfectly. For all that he believes, it might very well be the world he lives in. Tell him and show him all about it, claiming that it is the real world; he will never be at all surprised, unless it surprises him to find that he has turned out to be right in all his beliefs without exception. Nothing he believes – no propositional object of his belief – is false at this world.
> However, the proposition (actually) expressed by 'London is pretty' […] is false at this world.» (Lewis 1981b: 286)

Pierre tem uma conceção extremamente pobre de Londres. Basicamente, concebe-a como a cidade a que os seus vizinhos chamam 'Londres'. Este é o papel que Londres tem no mundo subjetivamente construído por ele. Tendo em conta que receber o nome 'Londres' de uma certa comunidade não é a essência de nenhuma coisa (provavelmente, nem sequer parte dessa essência), existem noutros mundos possíveis cidades diferentes de Londres que ocupam, relativamente a outros sujeitos, exatamente o mesmo papel que Londres ocupa para Pierre. Um desses sujeitos assim situados perante uma cidade, chamemos-lhe André, tem vizinhos que chamam 'Londres' a Bristol e é por meio de ter adquirido deles esse nome que ele pensa acerca de Bristol. Consegue Pierre distinguir a situação de André daquela em que ele mesmo se encontra? Talvez consiga. Imaginemos que André nunca estudou lógica. Pierre sabe perfeitamente que é um lógico (vamos assumir), e assim facilmente rejeita a situação de André como podendo ser aquela em que se encontra. Mas imaginemos, em vez disso, que as características de André se adequam totalmente às crenças *de se* de Pierre. Por exemplo, Pierre acredita que vive numa casa de madeira. André vive numa casa de madeira. Pierre acredita que se chama 'Pierre'. André

chama-se ‘Pierre’. Pierre acredita que a relva é azul. André vive num mundo em que a relva é azul. E por aí em diante. Assumimos que Pierre só tem crenças *de se* verdadeiras. Agora, consegue Pierre distinguir a situação dele e de André? Obviamente que não. Mesmo com tudo em que ele acredita, não pode excluir a hipótese de ser idêntico a André. Ora, se Pierre conseguisse identificar Londres entre as várias coisas possíveis, saberia perfeitamente que não é André. Ele saberia que a cidade de que André ouviu falar com o nome ‘Londres’ não é a mesma que ele ouviu falar com esse nome. Mas ele não é capaz de saber isso. O máximo que ele consegue é dizer que cidades se adequam ao papel que ele atribui subjetivamente a Londres – nada mais. E Bristol do mundo de André ocupa tão bem esse papel relativamente a André quanto Londres do mundo atual relativamente a Pierre.

E é por isso que Pierre não tem a crença *de dicto* de que Londres é bela. Uma crença *de dicto* restringe a classe de mundos que um sujeito aceita que podem ser o atual. Tendo a crença de que a neve é branca, por exemplo, um sujeito exclui a atualidade dos mundos em que a neve não é branca. Pierre não é capaz de identificar Londres nos vários mundos. Acabamos de ver que ele só é capaz de dizer que cidades satisfazem a conceção (extremamente pobre) que ele tem de Londres. Assim, um mundo em que Bristol satisfaz essa conceção e é bela, e em que Londres difere imensamente da atualidade e é feia é um mundo que Pierre aceita como podendo ser o atual. (Antes assumimos que a conceção total que Pierre tem de Londres é que esta é a cidade a que os vizinhos *dele* chamam ‘Londres’. Esta é uma conceção *de se*. Uma cidade satisfaz esta conceção apenas *relativamente a certos sujeitos*. O que significa então dizer que Londres ou Bristol satisfazem essa conceção num mundo? Talvez que satisfazem a condição de serem chamadas ‘Londres’ por uma comunidade qualquer nesse mundo.)

Conseguir identificar uma coisa em vários mundos é ter uma conceção da essência dessa coisa. Na teoria das contrapartes defendida por D. Lewis, a essência de uma coisa é a propriedade exemplificada por todas as contrapartes dessa coisa e nada mais. Tendo em conta que a relação de contraparte é uma relação comparativa de semelhança, conhecer a essência envolve (1) conhecer de uma maneira mais ou menos aprofundada as características atuais de uma coisa (algumas delas extrínsecas, como a origem) e (2) a

importância relativa das várias características.[71] Pierre está longe de conhecer estes aspetos de Londres, assim como, em geral, cada um de nós está longe de conhecer esses aspetos da maior parte das coisas. É claro que se Pierre tivesse uma conceção da essência de Londres, isso seria suficiente para ele ter crenças *de re* acerca dessa cidade. Em geral, podemos dizer então que quando (1) um sujeito X tem a crença *de dicto* de que uma coisa com a essência E tem a propriedade F e (2) Y é a coisa no mundo atual que tem E, é verdade que (3) X tem a crença *de re* de que Y tem a propriedade F. Pierre, porém, não conhece a essência de Londres, mas tem uma crença *de re* acerca de Londres. Em geral, não conhecemos a essência das coisas acerca das quais temos crenças *de re*. Mostramos assim que as crenças *de dicto* não subsumem as crenças *de re*.

A proposta de D. Lewis é que uma crença *de re* é um caso particular de atribuição de uma propriedade a um objeto *através de uma descrição*. Descrições estão envolvidas no conteúdo *de dicto* e *de se* e fazem com que esse conteúdo diga respeito a determinadas coisas nos vários mundos. Eu acredito que o professor de Alexandre Magno recebeu o nome 'Aristóteles'. Com esta crença *de dicto*, estou a atribuir a propriedade de ter recebido o nome 'Aristóteles' a Aristóteles, concebido através da descrição "o professor de Alexandre Magno". (A mesma crença num mundo em que o professor de Alexandre Magno foi Sócrates serve para atribuir uma propriedade a Sócrates através dessa

71 A relação de contraparte é qualitativa. Dois mundos com o mesmo carácter qualitativo não podem diferir quanto às partes de cada um que são contrapartes de coisas de um qualquer outro mundo. Mas isso pode acontecer com a relação de identidade entre coisas em vários mundos. Esta não é determinada unicamente pelo carácter qualitativo dos mundos em que as coisas se encontram. Chega-se a um ponto em que não podemos apelar a mais características dos objetos para determinar se eles são ou não idênticos. Possivelmente nesse ponto chegamos a um facto básico: este objeto é idêntico àquele, e esse outro semelhante não, simplesmente porque não, e nada há mais a dizer. Talvez essa relação nem seja determinada sequer por algum elemento qualitativo. Numa variedade extrema desta teoria, talvez qualquer coisa seja idêntica a algo com quaisquer características imagináveis num outro mundo distante. (Um homem atual seria idêntico a um ovo noutro mundo e a uma galinha noutro ainda.) Aparentemente, neste caso, saber a essência de uma coisa é mais fácil que na teoria das contrapartes. A essência é a identidade – e nada mais simples há que a identidade: qualquer coisa é idêntica a ela própria e a nada mais. Na verdade, no entanto, a situação é mais complexa. Para conseguir identificar uma coisa pelos vários mundos, temos de saber que coisa é aquela que estamos a considerar. Temos de conseguir identifica-la não só entre todas as que existem no mundo atual, mas também entre todas as que existem no espaço lógico. Haver mundos qualitativamente idênticos que diferem quanto à identidade vem ainda piorar a situação. Como é que podemos saber que este pato atual é idêntico àquele pato noutro mundo, em vez de um outro pato semelhante nesse mundo? Há um mundo idêntico ao nosso em que isso acontece. Havendo indiscernibilidade entre mundos, há indiscernibilidade entre as partes desses mundos. Mas na teoria das contrapartes isso não afeta o conhecimento da essência: coisas indiscerníveis, provavelmente, têm exatamente a mesma essência, como não acontece na teoria de identidade entre coisas em vários mundos.

descrição.) Descrições podem também ser egocêntricas. Eu acredito que o autor do livro que acabei de ler se chama 'Aristóteles'. Suponhamos que eu estive há pouco a ler Descartes. Através dessa crença *de se* eu atribuo a propriedade de chamar-se 'Aristóteles' a Descartes, concebido pela descrição egocêntrica "o autor do livro que acabei de ler".[72] Em geral, (1) X atribui a propriedade F a Y através da descrição Z se e só se (1) X acredita que a única coisa que satisfaz Z é F e (2) Y é a única coisa no mundo atual que satisfaz Z.

Nem todas as descrições estão aptas a fazer com que um sujeito tenha crenças *de re* acerca do único objeto que satisfaz a descrição. Este exemplo é utilizado em (Lewis 1979a: 539). Atribuir a propriedade de espionagem a um homem através da descrição "o espião mais baixo" claramente não é um caso de crença *de re* acerca desse homem. Acabamos há pouco de ver que as essências são descrições adequadas para esse efeito, mas que raramente são utilizadas por serem difíceis de capturar. O que torna então adequadas outras descrições para além das relativas às essências?

> «It will not be possible to say precisely which relations are suitable, since it is often quite vague whether some case should or should not count as an example of belief *de re*. The vagueness is partly resolved in context, but differently in different contexts. Still, I can at least say something about what tends to make a relation be a suitable description.
> […] It will help to have a collection of examples, uncontroversial or so I hope, of relationships in which belief *de re* is possible. I can have beliefs *de re* (1) about my acquaintances, present or absent; (2) about contemporary public figures prominent in the news; (3) about the famous dead who feature prominently in history; (4) about authors whose works I have read; (5) about strangers I am somehow tracing, such as the driver of the car ahead of me, or the spy I am about to catch because he has left so many legible traces; and (7) about myself.

72 Podemos entender uma descrição como uma expressão linguística de propriedades, como 'o primeiro rei de Portugal', ou de relações entre uma coisa e um sujeito, se a descrição for egocêntrica, como 'a pessoa para quem estou a olhar'. Neste contexto, D. Lewis considera que é conveniente tratar antes as descrições como aquilo que esses itens expressam – i. e., propriedades e relações. (Mais simplesmente ainda, podemos dizer que as descrições são sempre relações. A descrição expressa por 'o primeiro rei de Portugal' é a relação estabelecida entre um sujeito e uma coisa se e só se essa coisa for o primeiro rei de Portugal.) O motivo para isso é que dessa forma não estamos limitados às propriedades e relações que podem ser expressas em palavras. Alguém pode atribuir propriedades a uma pessoa identificando-a através da imagem visual que tem dela (Lewis 1979a: 538). Além disso, encontro ainda outra vantagem nesta estipulação. Às vezes queremos dizer que uma pessoa acredita que a única coisa que satisfaz esta ou aquela descrição é tal e tal. Entendendo as descrições como propriedades ou relações, evitamos estar a falar de crenças metalinguísticas acerca de expressões que talvez nem façam parte do vocabulário do sujeito considerado.

> What have these cases in common? To put a name to it: a *relation of acquaintance*. To make a little more precise: in each case, I and the one of whom I have beliefs *de re* are so related that there is an extensive causal dependence of my states upon his; and this causal dependence is of a sort apt for the reliable transmission of information.» (Lewis 1979a: 539-42)

Um cão está à minha frente. As várias possibilidades de movimento (ou pelo menos algumas dessas possibilidades) que ele tem disponíveis correspondem a potenciais alterações na minha experiência percetual. Movendo-se de uma dessas maneiras possíveis, o cão vai causar em mim uma experiência a partir da qual eu obtenho a informação de como ele se está a mover. Estabelece-se assim uma relação de dependência causal alargada entre os estados do cão e os meus estados mentais, e essa dependência é de um tipo que permite normalmente a transmissão fiável de informação. É deste modo que se estabelece entre mim e este cão uma relação de contacto.[73] A ideia de D. Lewis é que uma relação assim está envolvida nas descrições adequadas às crenças *de re* que não são essências.

Para constituir uma relação de contacto, uma dependência causal deve estar apta a transmitir informação correta, ainda que não tenha de conseguir fazer isso efetivamente em todos os casos. A experiência que tenho do cão diante de mim pode ser correta perante várias possibilidades de movimento, mas falhar se ele se mover para perto de um espelho que provoca uma ilusão de ótica. A relação de contacto não deixa de estar operacional por causa desta falha. Ler um livro escrito por um historiador da filosofia que fala de Espinosa é uma fonte fiável de informação sobre esse filósofo racionalista. E através dessa leitura alguém pode estabelecer uma relação de contacto – muito indireta, é verdade – com Espinosa. Essa relação não é perdida se houver um erro algures no livro.

Em geral, temos então que um sujeito X atribui a propriedade F a um objeto Y se e só se (1) X acredita que a única coisa que satisfaz a descrição Z no mundo atual tem a propriedade F, (2) Y é a única coisa que satisfaz Z no mundo atual, e (3) Z captura a

---

73 Não é suficiente que haja uma dependência causal, por muito alargada que seja, para ser estabelecida uma relação de contacto. Consideremos este exemplo de D. Lewis: «My life may be remarkably entangled with that of some stranger. I may have caught his germs time and again. His driving may have caused traffic jams that made me late to many important appointments. He may have caused many people to go to places where they happened to meet me. And so on. In short, maybe my life would have been very different but for his doings. None of this, by itself, makes it possible for me to have beliefs *de re* about this stranger.» (Lewis 1979a: 542)

essência de Y ou Z é uma relação de contacto. (Quando a descrição envolve uma relação, é uma crença *de se* que está na base da crença *de re*, enquanto que quando a descrição envolve uma essência, é uma crença *de dicto*.)[74, 75]

A essência de uma coisa é a propriedade exemplificada por todas as contrapartes dessa coisa e nada mais. É impossível que vários sujeitos atribuam uma propriedade a diferentes objetos (que não sejam contrapartes um do outro) através da mesma essência. O conteúdo das crenças *de re* que envolvem essências é, por isso, restrito, admitindo que também é restrito o da crença *de dicto* em que esta se baseia. O mesmo não acontece, porém, com as crenças *de re* que envolvem relações de contacto, que têm sempre um conteúdo lato. Chamamos-lhes crenças apenas no sentido derivado, que depende da maneira como usamos a linguagem comum. Aquilo que estabelece uma relação de contacto com alguém pode não ser idêntico àquilo que estabelece a mesma relação de contacto com outra pessoa. Eu considero bela a pedra que me está neste momento a provocar uma certa imagem visual. Certamente noutro mundo possível uma pedra muito semelhante à que estou a ver está a provocar exatamente a mesma imagem a uma outra pessoa. Mesmo que essa pessoa tenha um estado interno semelhante ao meu, estamos a ter uma crença *de re* acerca de pedras diferentes.

Esta abordagem ao fenómeno das crenças *de re* serve para analisar algumas frases acerca de crenças, ainda que esteja longe de cumprir a tarefa bastante mais complexa de ser uma análise completa e sistemática de todas as frases desse género. Mais

---

74 Também podemos atribuir relações a sequências de objetos. No caso da atribuição de uma relação binária a um par ordenado (A, B), temos o seguinte esquema. (1) X acredita que a única coisa que satisfaz a descrição Z_1 e a única coisa que coisa que satisfaz a descrição Z_2 estabelecem a relação R, (2) A é a única coisa que satisfaz Z_1 no mundo atual e B é a única coisa que satisfaz a Z_2 no mundo atual, e (3) Z_1 e Z_2 capturam uma essência ou envolvem relações de contacto. Facilmente se constroem os esquemas para relações com mais que dois argumentos.

É também possível atribuir propriedades ou relações utilizando as suas essências ou relações de contacto que um sujeito estabelece com elas. No caso de atribuirmos uma propriedade F nessas condições a um objeto A, temos um esquema mais complexo. (1) X acredita que o único objeto que satisfaz a descrição Z_1 tem a única propriedade que satisfaz a descrição Z_2, (2) A é o único objeto que satisfaz Z_1 no mundo atual e F é a única propriedade que satisfaz Z_2 no mundo atual, e (3) Z_1 e Z_2 capturam uma essência (a primeira a essência de um objeto, a segunda a essência de uma propriedade) ou envolvem relações de contacto. E assim por diante com as várias relações.

75 Quando X tem uma crença *de se*, tem uma crença *de re* acerca de X, que acontece através de uma relação de contacto muito simples e direta: a relação de identidade. Ao discutirmos se os objetos das crenças *de se* podem ser proposições, vimos que uma crença *de se* não pode estar baseada na essência do próprio sujeito.

concretamente, esta abordagem permite-nos explicar como é que podem ser verdadeiras diferentes atribuições de conteúdo de sujeitos psicologicamente idênticos. Um habitante na Terra tem crenças acerca da água, e uma cópia dele na Terra Gémea tem crenças acerca de XYZ. Como? Analisando estes casos como atribuições de crença *de re*, descobrimos que o que estamos a dizer do habitante da Terra é que (1) ele acredita que a substância natural com que estabelece uma certa relação de contacto não especificada é desta ou daquela maneira, e (2) essa substância é água. E descobrimos que o que estamos a dizer do habitante da Terra Gémea é que (1) ele também acredita que a substância natural que estabelece uma certa relação de contacto não especificada é desta ou daquela maneira, mas (2) que essa substância é XYZ. Há uma parte idêntica e outra parte diferente no que é afirmado dos dois sujeitos. É idêntica a parte que fala do conteúdo no sentido básico. É diferente a parte que envolve a relação do sujeito com o ambiente, e que é relevante para o conteúdo derivado. Quanto ao conteúdo básico, não especificamos qual é a relação de contacto envolvida que torna em cada caso a respetiva frase verdadeira. Há a possibilidade de ser a mesma em ambos os casos e, desse modo, há a possibilidade de os dois sujeitos serem idênticos no que diz respeito ao conteúdo básico. É o que supusemos que acontece de facto. O habitante da Terra acredita que a substância que existe nos mares e nos lagos, que provoca nele experiências de um certo tipo, e que ele e os outros chamam 'água' é de uma determinada maneira, e o mesmo acontece com a cópia dele na Terra Gémea. Haveria um problema para a teoria da mente de D. Lewis se eles diferissem nesse tipo de conteúdo, mas nada no exemplo sugere que isso acontece. Eles diferem apenas no conteúdo derivado. Não se pretende que esse conteúdo seja funcionalmente adequado e, por isso, não há problema em que ele seja lato. (Facilmente podemos modificar o que acabei de dizer para se adequar ao caso da artrite.)

Este método de análise permite também explicar como é que podemos atribuir proposições contraditórias às crenças de Pierre, dizendo ao mesmo tempo que ele não está envolvido em qualquer contradição. Pierre acredita que Londres é bela. Uma vez mais, analisando esta frase como uma atribuição de crença *de re*, isto significa que (1) Pierre acredita que a cidade com que estabelece uma certa relação de contacto não especificada é bela, e (2) essa cidade é Londres. Pierre acredita que Londres não é bela. Estamos agora

a dizer que (1) Pierre acredita que a cidade com que estabelece uma certa relação de contacto não é bela, e (2) essa cidade é Londres. Contrariamente ao que acontecia no exemplo anterior, é idêntico tudo o que é dito de Pierre com as duas frases. Em particular, é idêntica a parte que diz respeito à relação de Pierre com o ambiente externo. Quanto à outra parte, que se prende apenas com o conteúdo básico, no entanto, como nela não está especificada a relação de contacto que torna verdadeiras essas frases, deixamos em aberto a possibilidade de ser ou não a mesma em ambos os casos. Com isso deixamos em aberto a possibilidade de o conteúdo básico não ser o mesmo, como acontece de facto. Não sendo o mesmo, as crenças não são contraditórias no sentido básico, mas apenas no sentido derivado.

Através da investigação empírica vamos conseguindo chegar a uma conceção cada vez mais alargada e pormenorizada dos objetos e propriedades que compõem o mundo à nossa volta. Inicialmente, tínhamos uma conceção bastante superficial das várias partes da realidade, que incidia muito mais na maneira como essas coisas afetam a nossa experiência ou nos aspetos delas que se tornam para nós salientes em virtude de algum interesse pragmático, do que propriamente na sua natureza intrínseca ou em características que fazem parte da sua essência.

Enquanto vai sendo alargada e aprofundada, essa conceção começa a incluir informação acerca dos componentes que formam as coisas e a maneira como esses componentes estão estruturados. Deste modo, o papel que as coisas adquirem na representação do mundo que cada um de nós subjetivamente constrói torna-se cada vez mais rico, e com isso restringimos as possibilidades que satisfazem esse papel. Antes de alguém conhecer a composição química da água, ninguém era capaz de ter uma conceção da água que fosse capaz de a distinguir de substâncias superficialmente semelhantes como a imaginária XYZ da Terra Gémea. A situação torna-se muito diferente quando alguns humanos descobrem a identidade teórica entre a água e H2O.

Talvez tenhamos capturado a essência da água ao descobrir que é H2O. Mas talvez não. Para que essa descoberta empírica capture a essência da água é necessário que, ao mesmo tempo, tenhamos capturado a essência dos componentes hidrogénio e oxigénio. Sem isso, continuamos a estar inaptos a distinguir as possibilidades em que a água existe

daquelas em que existe uma substância formada por moléculas com uma estrutura semelhante às da água e com componentes chamados 'hidrogénio' e 'oxigénio', ainda que não sejam idênticos, respetivamente, ao hidrogénio e oxigénio. É improvável que alguma vez alguém tenha conseguido capturar a essência desses componentes, e é improvável que o consigamos fazer até que tenhamos descoberto as propriedades fundamentais que estão na base do mundo atual.

Ainda assim, parece que podemos ter a esperança de chegarmos a capturar a essência da água e de qualquer outra parte da realidade. Para que essa esperança seja realizada precisamos de conseguir identificar as propriedades e relações fundamentais exemplificadas pelas coisas mais básicas. Só desse modo é que podemos conhecer a essência tanto das coisas mais básicas como das somas mereológicas dessas coisas. De acordo com o que D. Lewis argumenta em "Ramseyan Humility" (2009), no entanto, isso nunca irá acontecer. Nesse artigo, explica que a única evidência que temos para identificar as propriedades fundamentais é o papel que elas ocupam – i. e., o modo como se relacionam entre si e com outras propriedades menos básicas – e que essa evidência é insuficiente para cumprir essa tarefa. Como ele diz, «[...] to the extent that we know of the properties of things only as role-occupants, we have not yet identified those properties. No amount of knowledge about what roles are occupied will tell us which properties occupy which roles.» (Lewis 2009: 204) Ora, existe a possibilidade de a propriedade que ocupa um certo papel não ser aquela que ocupa esse papel atualmente (pode ser ocupado por diferentes propriedades existentes no mundo atual ou outras, estranhas ao nosso mundo). Esta é uma consequência da ideia de que não existe uma conexão necessária entre coisas distintas e que, por isso, qualquer combinação de elementos de um mundo forma um outro mundo possível (*ver* secção 1.7). Como é que podemos distinguir qualquer uma dessas possibilidades da atualidade? É provável que isso seja impossível. Se a única evidência disponível é o papel que as propriedades ocupam, essa evidência não nos ajuda minimamente.

A conclusão não é que não conseguimos identificar qualquer propriedade das coisas. Essa conclusão seria não só extremamente estranha, mas também absurda, porque teria a consequência de que não conseguimos ter qualquer crença *de dicto* ou *de se*. Tendo

uma crença, distinguimos entre os mundos que consideramos poderem ser o atual daqueles que consideramos não poderem, ou entre os indivíduos que consideramos poderem ser nós mesmos e os que consideramos não poderem. Obviamente, é necessário identificar algumas propriedades para que através delas consigamos distinguir essas várias possibilidades.

Em vez disso, a conclusão é que não conseguimos identificar (1) as propriedades fundamentais (perfeitamente naturais), (2) as propriedades estruturais construídas a partir destas e (3) *algumas* propriedades obtidas verofuncionalmente a partir de propriedades fundamentais ou estruturais. (As propriedades fundamentais, perfeitamente naturais, são tratadas na secção 3.4.) Nem todas as propriedades assim construídas estão fora do nosso alcance, como explica D. Lewis:

> «There are some exceptional truth-functional compounds, however, to which Humility[76] does not apply. We saw that being composed of an F bearing R to a G might be any of eight different structural properties. We can identify the disjunction of all eight, even if we can't identify any one of its disjuncts. Likewise, even if we cannot identify the fundamental or structural property that actually occupies a certain role, we can identify the property of having whatever property it is that occupies that role (in the case in question).» (Lewis 2009: 215)

Antes desta passagem, D. Lewis argumentava que não conseguimos identificar a propriedade estrutural de ser composto por um F que estabelece a relação R com um G, sendo F, G e R fundamentais, se não soubermos que propriedades ou relações expressam 'F', 'G' e 'R', como D. Lewis entende que acontece efetivamente. Conseguimos, no entanto, identificar a propriedade disjuntiva de ter uma (qualquer) dessas propriedades que podem ser aquela que é expressa por 'a propriedade de ser composto por um F que estabelece a relação R com um G'. São propriedades como estas que estão envolvidas no conteúdo restrito. (Adiante, veremos mais exemplos de propriedades deste género.) Das restantes, irremediavelmente só podemos pensar de uma maneira derivada, em que o mundo externo ter de contribuir.

---

76 'Humildade' é o nome dado à tese que temos estado a considerar. A escolha do nome é óbvia. De acordo com esta tese, o nosso conhecimento das propriedades é irremediavelmente incompleto e, por isso mesmo, humilde.

Os exemplos da Terra Gémea e da artrite mostram que algumas atribuições de conteúdo podem ser verdadeiras de certos sujeitos e falsas de outros com exatamente as mesmas características funcionais. Acontece isto porque as representações na mente de cada sujeito estão conectadas a diferentes itens no ambiente externo. Não havendo conexões necessárias entre diferentes existentes, não devemos generalizar esses resultados a todas as atribuições de conteúdo? Não é verdade que qualquer representação num cérebro pode estar relacionada com diferentes coisas daquelas com que está relacionada uma outra representação, que ocupa o mesmo papel causal noutro cérebro?

Um proponente de uma teoria que explica o conteúdo através das relações que os estados cerebrais estabelecem com coisas ao seu redor teria que responder afirmativamente a estas questões (*ver* secção 3.2). De acordo com este esboço de teoria, uma atitude é a crença de que a neve é branca porque está de alguma maneira conectada causalmente com a neve e com a brancura. (Assim como a crença de que Heloísa ama Abelardo tem de estar conectada com Heloísa, Abelardo e o Amor. Mas esta conexão não é suficiente. A teoria tem ainda de explicar a diferença entre esta crença e a de que Abelardo ama Heloísa. Uma maneira de explicar isto passa por defender a ideia de que existe uma linguagem de pensamento em que as expressões mais básicas, que referem os objetos externos, formam outras mais complexas, incluindo frases que expressam certas proposições, através de algumas regras sintáticas e semânticas.)

A teoria de D. Lewis, pelo contrário, não faz o conteúdo depender das relações que as representações têm com os objetos externos. Enquanto que em qualquer variedade da teoria que estávamos a considerar, as relações de contacto assumem um estatuto de condição de possibilidade de representação, na teoria de D. Lewis essas relações, a aparecerem, aparecem apenas quando estão integradas na própria representação.

Recapitulando, de acordo com D. Lewis, uma certa atitude tem este ou aquele conteúdo em virtude do papel causal interno que ocupa. E, como vimos anteriormente, quando o conteúdo é lato e depende de mais fatores para além do papel causal, ele é entendido como um conteúdo derivado que tem por base um conteúdo restrito acerca das relações de contacto que um sujeito julga ter com coisas circundantes.

D. Lewis encontra problemas sérios com qualquer variedade da teoria que estávamos a considerar. Em primeiro lugar, esta permite apenas obter conteúdo lato e, desse modo, não é capaz de acomodar certas atribuições de conteúdo restrito que fazemos corretamente a muitos sujeitos. (Apresentamos antes o problema oposto para a teoria de D. Lewis – mais concretamente, que não acomodava as atribuições de conteúdo lato.) Nem todas as frases acerca de crenças são suscetíveis de variação no valor de verdade relativamente a sujeitos internamente semelhantes, ao contrário daquilo que pode ser sugerido pela existência de casos como o da Terra Gémea. É o que D. Lewis explica aqui:

> «We should not jump to the conclusion that just any belief sentence is susceptible to Twin Earth examples. Oscar thinks that square pegs don't fit round holes. I don´t think you can tell an even halfway convincing story of how Twoscar, just by being differently acquainted, fails to think so too. Oscar believes there's a famous seaside place called 'Blackpool'; so does differently acquainted Twoscar, though of course it may not be Blackpool – not *our* Blackpool that he has in mind. Oscar believes that the stuff he has heard of under the name 'water' falls from clouds. So does Twoscar – and so does Twoscar even if you alter not only his acquaintance with water, but his relations of acquaintance to other things as well. You know the recipe for Twin Earth examples. You can follow it in these cases too. But what you get falls flat even as an example of how content is sometimes wide, let alone as evidence that content is always wide.» (Lewis 1994: 424)

A atribuição a alguém da crença *de se* de que estabelece uma certa relação com uma coisa que tem estas ou aquelas características é um caso de atribuição de conteúdo restrito que faz parte da linguagem comum. X acredita que aquilo a que chama 'água' cai das nuvens, assim como X*, apesar de X estar em contacto com água e X* com XYZ. Da mesma maneira, X acredita que sofre de uma doença chamada 'artrite', assim como X*, apesar de X estar a falar de artrite com esse nome e X* de uma classe mais alargada de doenças que inclui também uma inflamação na coxa. D. Lewis apresenta ainda o exemplo da crença *de dicto* de que parafusos quadrados não encaixam em buracos redondos. Esta crença envolve apenas atributos que dizem respeito à forma ou estrutura geométrica de uma coisa, como as propriedades de ser quadrado e de ser redondo, e a relação de encaixar, e a propriedade funcional de ser um parafuso. (A propriedade de ser um buraco talvez caiba na primeira categoria.) Nenhum destes atributos envolve a especificação dos componentes de uma coisa e da estrutura interna em que esses componentes se encontram,

e capturam apenas aspetos superficiais das coisas, que são facilmente observáveis. Comparemos, por exemplo, a crença de que estou a olhar para um parafuso com a crença de que estou a olhar para um parafuso de cobre. A primeira claramente tem um conteúdo restrito, mas é provável que a última tenha um conteúdo lato. Conhecer a essência da propriedade de ser um parafuso é alguma coisa que, aparentemente, é bastante simples, mas o mesmo não se pode dizer da essência da propriedade de ser um parafuso de cobre. Conhecer a essência de ser um parafuso de cobre envolve conhecer como é que elementos simples compõem o cobre e de que maneira é que eles têm de estar estruturados. Claramente antes de haver conhecimento científico o conteúdo de uma crença acerca de um parafuso de cobre era um conteúdo lato. (As propriedades expressas por 'redondo', 'quadrado' e 'parafuso' pertencem, aparentemente, à classe de propriedades que não estão abrangidas pela tese da humildade. Elas são, por isso, altamente disjuntivas ou dizem respeito à ocupação de certos papeis.)

Em segundo lugar, uma teoria destas tem a consequência de que um cérebro numa cuba idêntico ao cérebro de qualquer um de nós, por estar a interagir com um ambiente muito diferente do nosso, não tem nada de psicologicamente em comum connosco (Lewis 1994: 424-5). A verdade é que um cérebro destes está conectado através de impulsos elétricos que afetam os seus recetores nervosos apenas com componentes do computador que a ele está ligado e que assim controla toda a sua experiência. Por isso ele não tem qualquer pensamento sobre água, madeira, mesas, cérebros e cubas, mas apenas sobre aquilo que o computador utiliza para simular essas coisas na experiência do cérebro (Putnam 1981: cap. 1). Todo o conteúdo lato atribuível a um cérebro numa cuba refere-se apenas a essa realidade criada artificialmente em computador. Em termos de conteúdo lato, não há qualquer semelhança entre o cérebro de um de nós e o idêntico cérebro numa cuba. Ainda assim, parece óbvio que existem aspetos que ambos têm em comum a nível mental. O cérebro numa cuba pensa que está a ver uma coisa a que chama 'árvore', e o mesmo acontece com o cérebro numa pessoa, apesar de apenas este último estar a ver realmente uma árvore. Outro exemplo citado por D. Lewis é o do homem do pântano, apresentado assim em (Davidson 1987):

> «Suppose lightning strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, The Swampman, moves exactly as I did; according to its nature, it departs the swamp, encounters and seems to recognize my friends, and appears to return their greetings in English. It moves into my house and seems to write articles on radical interpretation. No one can tell the difference.» (Davidson 1987: 443)

Como o homem do pântano apareceu no mundo desta maneira, ele não tem qualquer relação de contacto com alguma coisa. Havendo apenas conteúdo lato, não há qualquer semelhança mental entre esse homem e Donald Davidson, de quem é uma réplica perfeita. Mas isto é absurdo. O homem do pântano não estabelece relações de contacto com nada, mas certamente julga que sim, tal como D. Davidson julga dele próprio. A diferença é que um está certo e outro está errado.

Em terceiro lugar, e por fim, uma teoria que não admite a existência de conteúdo restrito não nos fornece os meios para caracterizar o que há de diferente entre crenças que dizem o mesmo acerca de certas partes da realidade com que estão associadas, apesar de ocuparem papeis causais distintos. O conteúdo lato é insensível à maneira como os objetos que nele estão incluídos se relacionam com as crenças que esse conteúdo está a caracterizar. É o conteúdo restrito, na teoria de D. Lewis, que permite capturar isso, ao incluir nele as relações de contacto que um sujeito pensa ter com certas coisas.

É por isso que o conteúdo lato não está apto a caracterizar adequadamente a parte do papel causal das atitudes que tem a ver com a racionalidade, como afirma D. Lewis nesta passagem:

> «The furniture of the *Lebenswelt* which presents us with our problems of decision and learning consists, in the first instance, of objects given *qua* objects of acquaintance, and individuated by acquaintance. That is a matter of narrow content. If you are lucky, and you're never wrong or uncertain about whether the thing you're R-acquainted with is or isn't the same thing you're R-acquainted with, then we can talk about your beliefs and desires entirely by terms of wide content. We can safely let things *simpliciter* stand in for things-*qua*-objects-acquaintance. But if you're not so lucky, that won't work.» (Lewis 1994: 429)

Distinguir entre o conteúdo restrito de crenças com o mesmo conteúdo lato é irrelevante quando um sujeito conhece a identidade entre os objetos com que estabelece

relações de contacto. Retomando um exemplo anterior, essa não é a situação em que Pierre se encontra. Suponhamos que Pierre regressa a França e que com ele traz as crenças que tinha enquanto vivia em Londres. Descrevendo a psicologia de Pierre recorrendo a conteúdo lato vamos dizer que, entre outras coisas, ele tem uma crença de que Londres é bela e uma outra de que Londres não é bela, e ainda que ele não tem a crença de que a cidade a que chama 'Londres' é a mesma que a cidade a que chama 'London'. Como vimos antes (*ver* secção 4.3), já nesta situação encontramos um problema, porque é próprio de um sujeito racional que ele consiga detetar contradições nas suas crenças, uma tarefa que Pierre não consegue de maneira nenhuma completar através de meios lógicos. Mas há mais. Suponhamos ainda que ele quer ir urgentemente visitar uma cidade bela e é confrontado com duas alternativas de escolha. Ele pode apanhar um de dois autocarros que vão para Londres. Um deles tem informações em francês onde é dito que irá para 'Londres'. O outro tem as mesmas informações em inglês e em vez de 'Londres' fala de 'London'. Assumindo que outros fatores (o custo de cada viagem, por exemplo) não influenciam Pierre nesta decisão, quem conhece mais aprofundadamente as suas crenças irá dizer que ele, fazendo uma escolha racional, vai entrar no autocarro com a informação em francês. Não é possível dizer isso, no entanto, com base unicamente no conteúdo lato. O que de facto aconteceu com Pierre foi algo deste género. Ele viu que o autocarro em francês ia para uma cidade chamada 'Londres'. Ele tem a crença de que a cidade chamada 'Londres' é bela e, a partir daí, inferiu que esse autocarro viajava para uma cidade bela. Como é que podemos descrever isto com conteúdo lato? Pierre viu que o autocarro em francês ia para uma cidade chamada 'Londres'. Ele tem a crença de que Londres é bela. A não ser que ele saiba ainda que 'Londres' refere Londres, não consegue inferir que o autocarro vai para uma cidade bela. (Estou a assumir que ele não tem informação adicional.) Através da inferência que Pierre fez, descobriu que a melhor maneira de levar a cabo a viagem para uma cidade bela era escolher o autocarro em francês, e por isso optou por entrar nesse autocarro. Esse raciocínio, descrito através do conteúdo lato, é errado.

Tendo em conta que o conteúdo restrito é, assumindo o que D. Lewis defende, muito mais relevante para uma descrição completa da vida mental de um sujeito que o

conteúdo lato, como é que explicamos a abundância de atribuições de conteúdo na linguagem comum que têm de ser analisadas recorrendo a conteúdo deste último género? Esta é uma explicação oferecida por D. Lewis:

> «Often we know a lot about which singular propositions someone believes in this wide and derivative way; but we know less about *how* – in virtue of just which self-ascriptions and relations of acquaintance – he believes those singular propositions. So it's no surprise to find that our ordinary-language belief sentences often seem to be ascription of wide content.» (Lewis 1994: 427)

Detalhes da vida mental de um sujeito – e, em particular, detalhes acerca da maneira como ele concebe as coisas ao seu redor – estão normalmente escondidos de um observador externo. Podemos dizer que não são aspetos completamente privados, tendo em conta que podem ser descobertos através de dados em princípio acessíveis a qualquer um. Mas são aspetos que escapam a um olhar superficial, e que só podem ser alcançados após uma longa investigação empírica – provavelmente inalcançável a qualquer humano – dos papeis causais ocupados pelos vários estados cerebrais. É por isso expectável que a linguagem comum encontre formas de falar das atitudes que não estejam restritas ao que acontece dentro da mente de cada um.

Relacionadas com isto estão algumas considerações feitas, antes, em *On the Plurality of Worlds* (1986b), nas quais D. Lewis associa o idioma do conteúdo lato a certas tarefas práticas, como o trabalho cooperativo:

> «[...] when we are interested less in the subject's psychology and more in his dealings with the things around him, as happens if we are interested in him as a partner in cooperative work and as a link in channels for information, then it is otherwise. The more he and we ascribe the same properties to the same individuals, the better we fare in trying to coordinate our efforts to influence those individuals. We learn from him by trying to ascribe the same properties to the same things that he does. We reach him by trying to get him to ascribe the same properties to things that we do. What matters is *agreement* about how things are; and we agree not when we think alike, but when we ascribe the same properties to the same things.» (Lewis 1986b: 59)

Cada um de nós constrói subjetivamente uma imagem diferente do mundo circundante. Nessa imagem encontram-se as várias descrições através das quais concebemos os vários objetos que se encontram no mundo. E, ainda dentro dessa imagem,

concebemos também a vida mental daqueles que nos rodeiam. É importante, em muitos casos, associar as propriedades que atribuímos a certos objetos com as propriedades que os outros atribuem a esses mesmos objetos, independentemente da maneira como os concebem. É o que acontece quando queremos aprender com outros ou trabalhar em conjunto com eles. Esta ideia complementa a explicação que anteriormente consideramos para o facto de serem abundantes as atribuições de conteúdo lato na linguagem comum.

Conclusão

O respeito pelo senso comum parece-me um traço importante da filosofia da mente que David Lewis desenvolveu ao longo do seu trabalho. É verdade que o contacto firme com a opinião popular sempre foi uma preocupação metodológica de D. Lewis, mas também é verdade que não o conseguimos encontrar com a mesma facilidade em todas as teorias que ele foi propondo. O confronto mais flagrante com o senso comum acontece com o realismo extremo acerca dos possíveis. A crença de que todas as nossas ficções não contraditórias fazem parte da realidade é, no mínimo, extravagante. Ainda assim, essa crença tem consequências filosoficamente interessantes e, exatamente por isso, é merecedora de atenção.

Na teoria da mente, no entanto, creio que encontramos um exemplo paradigmático de uma teoria comprometida com a simplicidade e elegância que se procuram na atividade filosófica e, ao mesmo tempo, com a conformidade ao senso comum que é exigida. D. Lewis tem uma atitude realista relativamente à existência da mente e à sua eficácia causal no mundo. Não há qualquer intenção de eliminar as entidades mentais, ou de desvalorizar de alguma maneira o discurso psicológico quotidiano – tudo isto feito com um forte pressuposto de que o materialismo é a ontologia correta para o mundo atual. Esse realismo leva D. Lewis a ter confiança na viabilidade da interpretação radical, e a opor-se fortemente à admissão de indeterminação no conteúdo mental e no significado linguístico. Há realmente alguma coisa em que acreditamos, que desejamos ou que temos intenção de fazer. E o nosso comportamento verbal é, de facto, significativo – há coisas que queremos dizer com as palavras que usamos.

O único momento em que encontro a teoria da mente de D. Lewis e o senso-comum a poderem estar a correr em direções contrárias é quando nega que conseguimos identificar os *qualia* da nossa experiência, e que Mary adquire nova informação acerca do mundo quando vê a cor vermelha pela primeira vez. Mas, mesmo aí, está presente a preocupação de acomodar da melhor maneira possível as opiniões comuns que parecem estar ameaçadas.

Além disso, a análise dos estados mentais através de descrições funcionais consegue acomodar a ideia comum de que os estados mentais são estados internos a cada

sujeito, sem ter de os encarar como uma realidade privada acessível unicamente através da introspeção. Os conceitos mentais dizem respeito a estados caracterizados apenas pela maneira como se relacionam com estímulos e comportamentos publicamente observáveis. De algum modo, a insistência em tratar os estados mentais como internos leva à insistência de que é o conteúdo restrito que é adequado para uma descrição correta, e completa, da vida mental das pessoas. As crenças e os desejos estão «dentro da cabeça», e é apenas por uma contingência da linguagem comum que chamamos atitudes a factos que dependem do ambiente externo.

Outro traço da teoria de D. Lewis que queria realçar é o tratamento da mente como logicamente anterior à linguagem, e o consequente tratamento da linguagem como um comportamento entre outros, igualmente racionalizado pelas atitudes dos sujeitos. Esta abordagem permite a D. Lewis teorizar acerca da mente independentemente, na maior parte dos casos, da teorização sobre a linguagem. Esta separação é notória, principalmente, na maneira como D. Lewis fala do conteúdo restrito e lato. Além disso, tendo uma análise da mente que não envolve a linguagem, há a possibilidade de analisar a linguagem começando pela mente – o que me parece uma das principais vantagens da filosofia da linguagem de D. Lewis.

Quero apenas referir que o realismo de D. Lewis também é verificável na suposição de que existem objetos intencionais que servem como conteúdo das atitudes. Esses objetos abstratos, estranhos e, para alguns, logicamente obscuros são tratados de uma maneira mais rigorosa como construções feitas a partir dos recursos ontológicos fornecidos pela teoria dos conjuntos e pelo realismo modal. D. Lewis consegue ao mesmo tempo manter o vocabulário que denota objetos como propriedades e proposições, mantendo uma teoria do mundo que, em última análise, pode ser compreendida como requerendo apenas um vocabulário admitido por uma lógica extensional. Conceitos modais e psicológicos, no final do dia, são reduzidos a conceitos extensionais, e ficamos apenas com o padrão de instanciação de propriedades e relações pelos objetos que habitam os vários mundos, e pela ontologia necessária à teoria dos conjuntos. Numa base tão parcimoniosa, uma teoria do mundo realista relativamente à mente e à modalidade é,

na minha opinião, um feito notável. Obviamente que isto tem a grande desvantagem de exigir uma ontologia extravagante que muitos não querem – nem conseguem – aceitar.

Espero que as páginas anteriores contenham uma apresentação adequada e útil das teses que compõem (ou que são relevantes para) a abordagem de D. Lewis à mente.

Referências bibliográficas

Adams, Robert (1970), "Theories of Actuality", *Noûs*, 8: 211-231.

Anscombe, G. E. M. (1957), *Intention*, Cambridge: Harvard University Press.

Block, Ned (1980), "Troubles with Functionalism", in Ned Block (ed.), *Readings in Philosophy of Psychology*, Volume I, Cambridge: Harvard University Press, pp. 268–305.

Braddon-Mitchell, David e Jackson, Frank (2007), *Philosophy of Mind and Cognition: An Introduction*, Blackwell Publishers.

Burge, Tyler (1979), "Individualism and the Mental", *Midwest Studies in Philosophy*, 4 (1): 73-122.

Chalmers, David (1996), *The Conscious Mind*, Oxford: Oxford University Press.

Davidson, Donald (1973), "Radical Interpretation", *Dialectica*, 27: 313-328.

Davidson, D. (1982), "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association*, 60 (3): 441-458.

Frege, Gottlob (1982), "Über Sinn und Bedeutung", *Zeitschrift für Philosophie Und Philosophische Kritik*, 100 (1): 25-50.

Gibbard, Allan (1975), "Contingent Identity", *Journal of Philosophical Logic*, 4: 187-221.

Hempel, Carl e Oppenheim, Paul (1948), "Studies in the Logic of Explanation", *Philosophy of Science*, 15: 135–175.

Jackson, Frank (1995), "What Mary Didn't Know", in Paul J. Moser e J. D. Trout (eds.), *Contemporary Materialism: A Reader*, Routledge, pp. 180-89.

Jackson, F. e Priest, Graham (2004), *Lewisian Themes: The Philosophy of David K. Lewis*, Oxford: Oxford University Press.

Kripke, Saul (1979), "A Puzzle about Belief", in A. Margalit (ed.), *Meaning and*

Use, Dordrecht: D. Reidel, pp. 238-283.

Kripke, S. (1980), *Naming and Necessity*, Cambridge, MA: Harvard University Press.

Lewis, David (1966), "An Argument for the Identity Theory", *Journal of Philosophy*, 63: 17–25.

Lewis, D. (1968), "Counterpart Theory and Quantified Modal Logic", *Journal of Philosophy*, 65: 115-126.

Lewis, D. (1969a), *Convention: A Philosophical Study*, Oxford: Blackwell Publishers.

Lewis, D. (1969b), Review of Capitan and Merrill (eds.), *Art, Mind, and Religion*, *Journal of Philosophy*, 66: 22–27.

Lewis, D. (1970a), "Anselm and Actuality", *Noûs*, 4: 175-188.

Lewis, D. (1970b), "General Semantics", *Synthese*, 22: 18–67.

Lewis, D. (1970c), "How to Define Theoretical Terms", *The Journal of Philosophy*, 67: 427-446.

Lewis, D. (1971), "Counterparts of Persons and Their Bodies", *Journal of Philosophy*, 68: 203-211.

Lewis, D. (1972), "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy*, 50: 249-258.

Lewis, D. (1973a), "Causation", *Journal of Philosophy*, 70: 556–567.

Lewis, D. (1973b), *Counterfactuals*, Oxford: Blackwell Publishers.

Lewis, D. (1974), "Radical Interpretation", *Synthese*, 23: 331–344.

Lewis, D. (1975), "Languages and Language", in Keith Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, Volume VII, Minneapolis: University of Minnesota Press, pp. 3–35.

Lewis, D. (1979a), "Attitudes *De Dicto* and *De Se*", *Philosophical Review*, 88: 513–

543.

Lewis, D. (1979b), “Scorekeeping in a Language Game”, *Journal of Philosophical Logic*, 8: 339-359.

Lewis, D. (1980a), “Index, Context and Content”, in Stig Kanger and Sven Öhman (eds.), *Philosophy and Grammar*, Dordrecht: Reidel, pp. 79–100

Lewis, D. (1980b), "Mad Pain and Martian Pain", in Ned Block (ed.), *Readings in Philosophy of Psychology*, Volume I, Cambridge: Harvard University Press, pp. 216–32.

Lewis, D. (1981a), “Causal Decision Theory”, *Australasian Journal of Philosophy*, 59: 5–30.

Lewis, D. (1981b), “What Puzzling Pierre Does Not Believe”, *Australasian Journal of Philosophy*, 59: 283-289.

Lewis, D. (1983a), “Individuation by Acquaintance and by Stipulation”, *Philosophical Review*, 92: 3–32.

Lewis, D. (1983b), “New Work for a Theory of Universals”, *Australasian Journal of Philosophy*, 61: 343-377.

Lewis, D. (1983c), *Philosophical Papers*, Volume I, Oxford: Oxford University Press.

Lewis, D. (1984), “Putnam’s Paradox”, *Australasian Journal of Philosophy*, 62: 221–236.

Lewis, D. (1986a), “Events”, in (1986c), pp. 241–269.

Lewis, D. (1986b), *On the Plurality of Worlds*, Oxford: Blackwell Publishers.

Lewis, D. (1986c), *Philosophical Papers*, Volume II, Oxford: Oxford University Press.

Lewis, D. (1988), "What Experience Teaches", *Proceedings of the Russellian Society*, University of Sydney, 13: 29–57.

Lewis, D. (1992a), Critical Notice of Armstrong, *A Combinatorial Theory of*

Possibility, *Australasian Journal of Philosop*hy, 70: 211–224.

Lewis, D. (1992b), "Meaning Without Use: *Australasian Journal of Philosophy*, 70: 106-110.

Lewis, D. (1993), "Mathematics is Megethology", *Philosophia Mathematica*, 3: 3-23.

Lewis, D. (1994), "Reduction of Mind", in Samuel Guttenplan (ed.), *A Companion to Philosophy of Mind*, Oxford: Blackwell Publishers, pp. 412–431.

Lewis, D. (1995), "Should a Materialist Believe in Qualia?", *Australasian Journal of Philosophy*, 73: 140–144.

Lewis, D. (1997), "Naming the Colours", *Australasian Journal of Philosophy*, 75: 325-342.

Lewis, D. (2009), "Ramseyan Humility", in David Braddon-Mitchell and Robert Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*, Cambridge: MIT Press, pp. 203–222.

Lewis, Stephanie (2015), "Intellectual Biography of David Lewis (1941-2001)", in Loewer and Schaffer (eds.), *A Companion to David Lewis*, Wiley Blackwell, pp. 3-14.

McDaniel, Kris (2004), "Modal Realism with Overlap", in (Jackson e Priest (eds.) 2004), pp. 140-155.

Nagel, Thomas (1974), "What Is It Like to Be a Bat?", *Philosophical Review*, 83: 435-450.

Nemirow, Laurence (1990), "Physicalism and the Cognitive Role of Acquaintance", in William G. Lycan (ed.), *Mind and Cognition*, Blackwell, pp. 490-499.

Nolan, Daniel (2005), *David Lewis*, Chesham: Acumen Publishing.

Nozick, Robert (1969), "Newcomb's Problem and Two Principles of Choice", in Nicholas Rescher (ed.), *Essays in Honor of Carl G. Hempel*, pp. 114–146, Dordrecht: Reidel.

Plantinga, Alvin (1974), *The Nature of Necessity*, Oxford: Oxford University Press.

Putnam, Hilary (1973), "Reference and Meaning", *Journal of Philosophy*, 70: 699-711.

Putnam, H. (1975), "The Meaning of 'Meaning'", *Minnesota Studies in the Philosophy of Science*, 7: 131-193.

Putnam, H. (1977), "Realism and Reason", *Proceedings and Addresses of the American Philosophical Association*, 50: 483-498.

Putnam, H. (1980), "The Nature of Mental States", in Ned Block (ed.), *Readings in Philosophy of Psychology*, Volume I, Cambridge: Harvard University Press, pp. 223–231.

Putnam, H. (1981), *Reason, Truth and History*, Cambridge: Cambridge University Press.

Quine, Willard Van Orman (1953), "On What There Is", in *From a Logical Point of View*, Cambridge, MA: Harvard University Press.

Quine, W. V. O. (1960), *Word and Object*, Cambridge, MA: MIT Press.

Rosen, Gideon (2015), "On the Nature of Certain Philosophical Entities: Set Theoretic Constructionalism in the Metaphysics of David Lewis", in in Loewer and Schaffer (eds.), *A Companion to David Lewis*, Wiley Blackwell, pp. 382-398.

Schwarz, Wolfgang (2015), "Analytic Functionalism", in Loewer and Schaffer (eds.), *A Companion to David Lewis*, Wiley Blackwell, pp. 504-518.

Smart, J. J. C. (1959), "Sensations and Brain Processes", *The Philosophical Review*, 68: 141-156.

Soames, Scott (2015), "David Lewis's Place in Analytic Philosophy", in Loewer and Schaffer (eds.), *A Companion to David Lewis*, Wiley Blackwell, pp. 80-98.

van Inwagen, Peter (1986), "Two Concepts of Possible Worlds", in *Midwest Studies in Philosophy XI*, T. Uchling, and H. Wettstein (eds.), Minneapolis, University of Minnesota Press, pp. 185-213.

Williams, J. R. G. (2015), "Lewis on Reference and Elegibility", in Loewer and

Schaffer (eds.), *A Companion to David Lewis*, Wiley Blackwell, pp. 367-381.

Anexo 1

O método de definição dos termos teóricos apresentado em (Lewis 1970c) envolve alguns mecanismos formais que permitem evitar a circularidade viciosa quando uma teoria introduz mais do que um termo. Faço agora a uma apresentação mais formal e completa desse método e dos mecanismos que utiliza (*ver* secção 2.1).

Sejam t_1, ..., t_n os termos introduzidos por uma teoria T (os termos-T).[77] Começamos por tratar T como uma única frase, o *postulado* de T, formando a conjunção das frases axiomáticas de T:

(1) $T(t_1, ..., t_n)$.

É conveniente tratar todos os termos teóricos que aparecem em T como nomes, como é feito em (Lewis 1970c: 429, Lewis 1972: 253). No entanto, predicados e functores também podem ser termos teóricos, e por isso é necessário encontrar uma maneira de acomodar expressões dessas categorias ao esquema desenhado para lidar com nomes. Para isso, substituímos todas as frases em que ocorre um predicado teórico com n lugares de argumento, $F(x)$, $F(x, y)$, $F(x, y, z)$, …, por frases como $E(\eta, x)$, $E(\eta, x, y)$, $E(\eta, x, y, z)$, …, que contêm um predicado com n+1 lugares de argumento e em que o primeiro desses lugares é ocupado com um nome para a propriedade que é expressa por F. O predicado F deve ser verdadeiro de uma sequência $(\eta, \alpha_1, \ldots, \alpha_n)$ se e só se o predicado F for verdadeiro da sequência $(\alpha_1, \ldots, \alpha_n)$. Assim, as frases de cada um dos pares $F(x)$ e $E(\alpha, x)$, $F(x, y)$ e $E(\alpha, x, y)$, $F(x, y, z)$ e $E(\alpha, x, y, z)$, …, têm exatamente as mesmas condições de verdade.

Podemos usar um processo semelhante se o termo teórico for um functor f. Substituímos todas os termos em que f ocorre, $f(x)$, $f(x, y)$, $f(x, y, z)$, …, por expressões como $e(ß, x)$, $e(ß, x, y)$ e $e(ß, x, y, z)$, …, nas quais ocorre um functor e e um nome $ß$ para uma função que ocupa o primeiro lugar de argumento. Os valores da função e vão

[77] Seguindo a notação de D. Lewis, cada termo-T é representados pela letra t indexada por um número natural que o distingue dos restantes. Associando arbitrariamente o número natural i a um termo teórico, podemos dizer, por conveniência, que esse termo teórico ocupa o lugar i da sequência de n termos teóricos que ocorre em T, sendo que $i \geq 0$ e $i \leq n$.

corresponder aos valores de f do seguinte modo: para qualquer $x, y, \ldots$, e qualquer função $ß$, $e(ß, x, y, \ldots) = f(x, y, ...)$.

Agora, modificando (1), podemos obter a *frase de realização* de T, substituindo cada termo teórico por uma variável.[78] Estas variáveis vão ocorrer livres e, por isso, a frase de realização é uma frase aberta que pode ser satisfeita por certas sequências de objetos.

(2) $T(x_1, \ldots, x_n)$.

Dizemos que realiza T qualquer sequência de objetos que satisfaz (2). A *frase de Ramsey* de T afirma que pelo menos uma sequência faz isso. Esta frase é construída através da colocação de quantificadores existenciais que ligam cada uma das variáveis que aparecem livres em (2):

(3) $\exists x_1, \ldots, \exists x_n \; T(x_1, \ldots, x_n)$.

Agora, formamos a *frase de Carnap* de T, que é basicamente uma frase condicional que tem (2) como antecedente e (1) como consequente:

(4) $\exists x_1, \ldots, \exists x_n \; T(x_1, \ldots, x_n) \rightarrow T(t_1, \ldots, t_n)$.

A conjunção de (3) e (4) é logicamente equivalente a (1). Por isso, têm como consequência exatamente as mesmas frases – nomeadamente, os teoremas de T.[79] De entre estas, podemos separar as frases que apenas contêm termos antigos (as frases-A) e as que contêm pelo menos um termo teórico (frases-T). Chamamos K ao conjunto de

[78] É necessário que a cada variável corresponda apenas um dos termos teóricos e que a mesma variável ocupe todos os lugares que o termo teórico correspondente ocupava. Por conveniência, representamos todas as variáveis com uma letra apenas, x, acompanhada de um número natural, o qual deve ser o mesmo, em cada variável, que o número natural que aparece no termo teórico correspondente.

[79] Da frase de Ramsey e da frase de Carnap podemos derivar o postulado por *modus ponens*. A partir do postulado podemos obter a frase de Ramsey por generalização existencial e, como a partir de uma frase ϕ podemos obter uma condicional que tenha ϕ como consequente, podemos obter a frase de Carnap a partir do postulado (Lewis 2009: 220, nota 7).

todos os teoremas de T, K(A) ao subconjunto de K que contém os teoremas que são frases-A, e K(T) ao subconjunto de K que contém os teoremas que são frases-T.

De (3) isoladamente seguem-se todas as frases que são elementos de K(A).[80] Assim, as únicas frases que pertencem a K e não são consequência de (3) pertencem a K(T).

De (4) isoladamente seguem-se elementos tanto de K(A) como de K(T). Nenhum deles, ainda assim, possui qualquer conteúdo factual.[81] Tendo em conta que de nenhuma frase sem conteúdo factual se seguem outras com conteúdo desse género, temos então que a contribuição de (4) para a derivação de frases em K a partir da conjunção de (3) e (4) não é um acréscimo em conteúdo factual. Assim, o conteúdo factual presente nas frases de K(T) tem de estar presente em K(A), querendo isto dizer que os termos teóricos são dispensáveis para aquilo que T tem a dizer acerca de questões de facto.

O enunciado (4) estabelece condições suficientes para a verdade de (1). É afirmado por (4) que se existe uma sequência de objetos que realiza T, então (1) é verdadeira. Fazendo isto, (4) está a interpretar parcialmente os termos teóricos que aparecem em T, de uma maneira que veremos já de seguida.

Se (3) for falsa, (4) nada tem a dizer acerca da interpretação dos termos-T. Nesse caso, (4) permanece verdadeira independentemente da referência dos termos-T. É por esse motivo que (4) não é capaz de fornecer uma interpretação total desses termos.

Contudo, se (3) é verdadeira, (4) passa a ser informativa acerca da interpretação dos termos-T. Há que distinguir, ainda assim, o caso em que (3) é verdadeira havendo apenas uma realização de T daquele em (3) é verdadeira havendo mais do que uma.

---

80 A prova desta afirmação, apresentada em (Lewis 2009: 219, nota 6), é a seguinte. Seja S uma qualquer frase-A implicada pelo postulado:

(1) Então [T(*t*) só se S] é uma verdade lógica.

(2) Daqui se segue que $\forall x$ [T(*x*) só se S] é uma verdade lógica.

(3) Então [$\exists x$ (T(*x*)) só se S] é uma verdade lógica.

(4) Daqui se conclui que da frase de Ramsey se segue S.

D. Lewis afirma que o passo de (1) para (2) é uma aplicação particular da regra de generalização universal, que apesar de não ser válida para qualquer termo singular, é válida quando estamos a lidar com verdades lógicas.

81 Neste contexto, quando digo que uma frase tem conteúdo factual quero dizer que dela se segue uma frase existencial, uma frase cuja verdade dependa da existência de um objeto num certo domínio.

No primeiro caso, (4) interpreta satisfatoriamente os termos-T. Tendo em conta que a sequência de objetos nomeados pelos termos-T é uma realização de T, e que não há mais nenhuma realização para além dessa, concluímos que os termos-T nomeiam os objetos que pertencem à sequência que realiza T. (Mais precisamente ainda, o termo que ocupa o lugar *i* da sequência de termos, t_i, nomeia o objeto que ocupa o lugar *i* da sequência de *n* elementos que realiza T, sendo $i \geq 0$ e $i \leq n$.)

No segundo caso, em que há realização múltipla de T, (4) afirma apenas que os termos teóricos nomeiam os objetos que pertencem a alguma das sequências que realizam T, não conseguindo distinguir uma de entre essas. D. Lewis teve várias opiniões, ao longo do seu trabalho, acerca do que acontece à referência dos termos-T nesta situação. De acordo com a proposta mais antiga (Lewis 1970c: 432-33, Lewis 1972: 252), a realização múltipla de T torna vazia a referência dos termos teóricos. Acontecendo isto, a frase de Ramsey permanece verdadeira, mas o postulado de T é falso – e, consequentemente, é falsa também a frase de Carnap, e inútil relativamente à interpretação dos termos. Esta proposta dependia da suposição de que T implicitamente afirma que possui apenas uma realização para que, desse modo, conseguisse interpretar satisfatoriamente os termos que introduz. De facto, para que as definições explícitas dos termos teóricos se sigam de T, é necessário que a teoria elimine a possibilidade de realização múltipla. Como veremos adiante, assumindo isto, T implica uma série de afirmações de identidade envolvendo os termos teóricos e as descrições definidas que os definem.

Não podemos dizer o mesmo se assumirmos a posição que D. Lewis adotou mais recentemente em (Lewis 1994: 417, 1997: 334), de acordo com a qual, em caso de realização múltipla, os termos teóricos referem indeterminadamente os correspondentes elementos das várias sequências que realizam T. Assim, o postulado de T passa a ser verdadeiro em pelo menos algumas das interpretações do termos-T.[82] A frase de Carnap passa também a ser verdadeira em algumas interpretações. Nenhuma definição explícita

82 Em algumas interpretações, e não todas, pela seguinte razão. Se duas sequências sem elementos em comum, (α_1, …, α_n) e (β_1, …, β_n), realizam T, cada termo-T tem duas interpretações possíveis: t_1 designa ambiguamente α_1 e β_1, …, e t_n designa ambiguamente α_n e β_n.. Se, ao considerarmos a verdade de T, optarmos por uma desambiguação dos termos que atribua a alguns objetos da primeira sequência e a outros objetos da segunda, T será falsa.

dos termos, contudo, pode ser obtida logicamente nestas circunstâncias a partir do postulado de T.

Para termos as vantagens de ambas as abordagens – ou seja, termos a possibilidade de chegar a definições explícitas dos termos-T e permitir que estes não tenham uma referência vazia em caso de realização múltipla – podemos escrever explicitamente no postulado de T que este não admite realização múltipla e, de seguida, permitir que as descrições definidas *impróprias* fornecidas pelas definições explícitas que a partir daí forem obtidas tenham um referente indeterminado se forem descrições.

Temos, assim, uma versão fraca e uma versão forte do postulado de T, não admitindo esta última a possibilidade de realização múltipla. Para chegar às definições explícitas, começamos então por escrever essa versão mais forte do postulado:

$$(1^*)\ \mathrm{T}(t_1, \ldots, t_n) \wedge \forall x_1, \ldots, \forall x_n\, \mathrm{T}(x_1, \ldots, x_n) \rightarrow (x_1 = t_1 \wedge \ldots \wedge x_n = t_n)).$$

A partir do postulado assim ampliado, seguem-se duas novas frases: os *postulados de significado*. Um desses postulados especifica a denotação dos termos-T no caso em que T é unicamente realizada. O outro afirma que se não houver uma única realização de T – mais concretamente, se não houver qualquer realização de T ou se houver mais do que uma – os termos ficam sem denotação (Lewis 1970c: 434-35, 1972: 254).

O primeiro postulado pode ser obtido substituindo o antecedente da frase de Carnap por uma modificação da frase de Ramsey, de modo a admitir apenas uma sequência a realizar *T*. Escrevemos assim esta frase de Ramsey modificada:

$$(3^*)\ \exists x_1, \ldots, \exists x_n\ (\mathrm{T}(x_1, \ldots, x_n) \wedge \forall y_1, \ldots, \forall y_n\, \mathrm{T}(y_1, \ldots, y_n) \rightarrow (x_1 = y_1 \wedge \ldots \wedge x_n = y_n)).$$

Daqui em diante, por razões de conveniência, esta frase aparecerá abreviada assim:

$$(3^*)\ \exists_1 x_1, \ldots, \exists_1 x_n\ \mathrm{T}(x_1, \ldots, x_n).$$

Obtemos, agora, a frase de Carnap modificada:

(4*) $\exists_1 x_1, \ldots, \exists_1 x_n \; T(x_1, \ldots, x_n) \rightarrow T(t_1, \ldots, t_n)$.

O segundo postulado pode ser algo deste género:

(5) $\sim \exists_1 x_1, \ldots, \exists_1 x_n \; T(x_1, \ldots, x_n) \rightarrow \sim \exists x \, (x = t_1 \vee \ldots \vee x = t_n)$.[83]

A relação lógica entre (1*) e a conjunção de (3*) e (4*) é análoga àquela que era estabelecida entre (1) e a conjunção de (3) e (4). Mais especificamente, todas as frases-A implicadas por (1*) são implicadas também por (3*). As restantes frases que são consequência de (1*), mas não da (3*) isoladamente, são obtidas pela conjunção de (3*) com (4*). Todo o conteúdo factual de T é expresso pela frase de Ramsey modificada.

Através dos postulados de significado encontramos, finalmente, um modo de expressar a componente semântica ou analítica de T que permite interpretar os termos-T. A componente factual ou existencial é fornecida pela frase de Ramsey convencional ou modificada, conforme pretendamos ou não admitir a realização múltipla.

Os postulados do significado, em conjunto, são equivalentes a uma série de afirmações de identidade que contêm cada uma, de um lado, um termo teórico, e, do outro, uma descrição definida, obtida de uma maneira simples a partir da frase modificada de Ramsey, que interpreta esse termo. Chegamos assim às definições explícitas dos termos teóricos:

(6.1) $t_1 = \iota x_1 : \exists_1 x_2, \ldots, \exists_1 x_n \; T(x_1, x_2 \ldots, x_n)$.

…

83 A ideia que se pretende expressar com o consequente é de facto metalinguística – mais concretamente, que não existe denotação dos termos teóricos. Para que este consequente seja verdadeiro temos de admitir na linguagem termos sem denotação. Uma lógica adequada neste contexto deve satisfazer as condições explicitadas na nota 82.

(6.*n*) $t_n = \iota x_n : \exists_1 x_1, \ldots, \exists_1 x_{n-1}\ T(x_1, \ldots, x_{n-1}, x_n)$.[84]

(Lewis 1970c: 437-38)

Cada uma destas afirmações diz-nos que o objeto nomeado por um certo termo teórico é aquele objeto, seja ele qual for, que ocupa um certo papel teórico especificado em T: que ele tem estas ou aquelas características, e relaciona-se desta ou daquela maneira com outros objetos nomeados unicamente através dos termos antigos que aparecem em T.

Os termos teóricos não ocorrem em nenhuma das descrições definidas que compõem as afirmações anteriores. Deste modo, não há qualquer circularidade em definir cada um dos termos recorrendo a uma destas descrições.

D. Lewis admite que mesmo os membros de uma sequência que realiza imperfeitamente T servem como referentes dos termos teóricos introduzidos por T, não havendo outra que realize mais adequadamente a teoria (Lewis 1970c: 432, Lewis 1972: 252). Podemos dizer que uma sequência assim realiza T\F: a diferença entre o conjunto total de frases de T e o subconjunto F de frases de T que não são satisfeitas pela sequência. (Nesta operação, talvez sejam eliminadas todas as frases que contêm determinados termos teóricos, e por isso podemos ter sequências de menor comprimento a realizar esse subconjunto de T.)

Podemos codificar esta liberalização nos próprios postulados do significado, assim como é proposto em (Schwarz 2015: 506). Primeiramente, formam-se frases de Carnap tendo como antecedentes versões crescentemente mais fracas do postulado de T:[85]

---

84 Como nota D. Lewis, para que estas afirmações de identidade sejam equivalentes aos postulados de significado, é necessário que sejam verdadeiras mesmo no caso em que não exista algo que satisfaça as descrições ou que seja nomeado pelos termos teóricos (basicamente, a situação em que o segundo postulado de significado é não trivialmente verdadeiro). Assim, é necessário rejeitar, pelo menos para este caso, a análise russelliana das descrições definidas, segundo a qual a falta de satisfação da descrição levaria a frase a ser falsa, e uma abordagem fregeana aos termos sem referência, que invalidaria a atribuição de um valor de verdade às frases. D. Lewis propõe que devemos assumir, para este caso, uma teoria dos termos sem denotação que permita tomar como verdade uma afirmação de identidade com dois termos sem denotação e como falsa no caso em que apenas um dos termos não tem denotação (Lewis 1970c: 430, Lewis 1972: 254, nota 11).

85 Sigo aqui a apresentação de W. Schwarz, utilizando apenas um termo teórico e uma variável. O mecanismo é facilmente expandido para qualquer número pretendido de termos e variáveis.

(7.1) $\exists x\, \mathrm{T}(x) \rightarrow \mathrm{T}(t)$.

(7.2) $(\sim \exists x\, \mathrm{T}(x) \wedge \exists x\, \mathrm{T}'(x)) \rightarrow \mathrm{T}'(t)$.

(7.3) $(\sim \exists x\, \mathrm{T}(x) \wedge \sim \exists x\, \mathrm{T}'(x) \wedge \exists x\, \mathrm{T}''(x)) \rightarrow \mathrm{T}''(t)$.[86]

…

Agora, seja a frase T*(x) uma abreviatura de:

(8) $\exists x\, \mathrm{T}(x) \vee (\sim \exists x\, \mathrm{T}(x) \wedge \exists x\, \mathrm{T}'(x)) \vee (\sim \exists x\, \mathrm{T}(x) \wedge \sim \exists x\, \mathrm{T}'(x) \wedge \exists x\, T''(x)) \vee$ ....

A conjunção de (7.1), (7.2), (7.3), e por aí em diante, é equivalente à seguinte frase de Carnap:

(9) $\exists x\, \mathrm{T}^*(x) \rightarrow \mathrm{T}^*(t)$.

Para fazer coincidir esta proposta de Wolfgang Schwarz com a de D. Lewis, teríamos de substituir (9) pela correspondente frase de Carnap modificada. E, por fim, teríamos de aplicar o mesmo procedimento para escrever o segundo postulado do significado:

(10) $\sim \exists x\, \mathrm{T}^*(x) \rightarrow \sim \exists x\, (x = t)$.[87]

86 T´e T´´são modificações de T análogas a T\F.

87 Aplica-se ao consequente deste enunciado o mesmo que se disse na nota 81 aplicar-se ao consequente do segundo postulado do significado.

Anexo 2

Frank Jackson apresentou um forte argumento contra qualquer forma de materialismo, normalmente designado como *argumento do conhecimento* (Jackson 1995). Mary conhece tudo o que há para conhecer acerca das propriedades físicas instanciadas no presente, passado e futuro do mundo possível atual, e tudo aquilo que a partir daí é superveniente – incluindo assim os papeis causais dos vários estados físicos, simples ou complexos. Mas Mary aprendeu tudo isto sem alguma vez ter visto cores diferentes do preto e do branco. Durante todo esse tempo, viveu num quarto com paredes brancas e pretas, e toda a sua aprendizagem foi feita através de transmissões televisivas, igualmente a preto e branco. Um dia, Mary liberta-se do local onde esteve toda a sua vida e pela primeira vez tem uma experiência da cor vermelha. O argumento de F. Jackson consiste em dizer que:

(1) Se o materialismo estiver correto, enquanto Mary estava no quarto sem cores já sabia tudo o que havia para saber sobre o mundo atual (@). O conhecimento de que dispunha era *completo*. Se assim não fosse, existiria um mundo possível *w* relativamente ao qual o conhecimento de Mary é completo, e a diferença entre @ e *w* teria de consistir em algum aspeto de @ que não é tratado pela física.

(2) No entanto, Mary não conhecia tudo sobre o mundo atual. Mais concretamente, não conhecia parte do seu aspeto fenomenal – aquele que diz respeito à experiência de vermelho – ou, dito de um outro modo, não conhecia os *qualia* de várias experiências, incluindo a de vermelho. Parte dessa informação só a obteve quando pela primeira vez teve a experiência de vermelho. Antes, existiam inúmeros mundos que Mary não podia excluir como sendo o seu. Todos eles eram idênticos a @ em aspetos físicos, mas entre eles existiam diferenças a nível fenomenal ou relativamente aos *qualia*.

Por fim, conclui-se, por *modus tollens*, que o materialismo é falso.

(Outro exemplo que podia servir para montar um argumento semelhante é apresentado em (Nagel 1974). A maioria dos morcegos tem um sistema percetivo radicalmente diferente do dos seres humanos. Essa diferença pode ser estudada completamente a nível neurofisiológico. No entanto, mesmo com toda essa informação, talvez nunca conseguiremos saber *como é* ser um morcego; ou seja, nunca poderemos

conhecer o aspeto fenomenal ou o *quale* da experiência do morcego. Então, não sabemos tudo o que há para saber sobre o morcego – e, *a fortiori*, sobre o mundo –, mesmo que saibamos tudo sobre a neurofisiologia do morcego.)

A resposta de D. Lewis a este argumento vai consistir em rejeitar (2). Não é verdade que Mary passa a saber *como é* ver a cor vermelha apenas na altura em que escapa do seu quarto? Sem dúvida, responde D. Lewis. Antes, não o sabia, assim como *nós* não sabemos como é ser um morcego. «Experience is the best teacher», afirma, «having an experience is the best way or perhaps the only way, of coming to know what that experience is like. No amount of scientific information about the stimuli that produce that experience and the process that goes on in you when you have that experience will enable you to know what it's like to have the experience.» (Lewis 1988: 78) Mas saber como é ver a cor vermelha ou saber como é ser um morcego não é, na sua perspetiva, adquirir informação sobre o mundo. Não consiste em colocar de lado certas possibilidades que antes estavam em aberto. Em vez disso, em consonância com Laurence Nemirow (Nemirow 1995), propõe que «knowing what it's like is the possession of abilities: abilities to recognize, abilities to imagine, abilities to predict one's behavior by means of imaginative experiments.» (Lewis 1983c: 131) Ou seja, ao ver a cor vermelha pela primeira vez, Mary passou a conseguir recordar e imaginar experiências desse tipo, e passou a poder usar essa habilidade para outras tarefas cognitivas. Aquilo que talvez nunca consigamos – nós, humanos – é ter uma estrutura neurofisiológica próxima dos morcegos que nos permita imaginar experiências do tipo das que esses animais têm. Deste modo, nunca poderemos saber como é ser um morcego.

Anexo 3

Tendo em conta as vantagens teóricas oferecidas pela distinção entre as propriedades mais e as menos naturais, D. Lewis considera uma das alterantivas adequadas a de tratar essa distinção como primitiva. Nesse caso, introduz-se o predicado '*x* é natural', aplicável a algumas propriedades e relações, sem que este receba qualquer análise (Lewis 1983b: 347, 1986b: 63-4). Neste sentido, a teoria das propriedades naturais pode continuar a ser uma teoria que requer apenas uma ontologia nominalista (exceção feita aos conjuntos, que são as únicas entidades abstratas – ou não concretas, ou não particulares - até agora admitidas), tal como a teoria das propriedades em geral.

De um outro modo, D. Lewis indica também que uma teoria deste género pode ir um pouco mais longe, continuando a manter o espírito nominalista, tomando como primitiva uma relação complexa de semelhança R entre objetos, que pode assim ser descrita:

$x_1, x_2, \ldots$ assemelham-se entre si e não se assemelham do mesmo modo a nenhum dos $y_1, y_2, \ldots$.

A ideia é que $x_1, x_2, \ldots$ sejam instâncias de uma única propriedade natural da qual nenhum dos $y_1, y_2, \ldots$ é uma instância. Agora, a partir de R, define-se a seguinte relação, chamemos-lhe N:

$$\exists y_1, y_2, \ldots \forall z \, R(z, x_1, x_2, \ldots, y_1, y_2, \ldots) \leftrightarrow (z = x_1 \vee z = x_2 \vee \ldots).$$

(Lewis 1983b: 347-48)

As propriedades naturais são, assim, aquelas que têm como elementos objetos que podem formar uma sequência $(a_1, a_2, \ldots)$ que satisfazem a relação N.

D. Lewis considera ainda mais duas alternativas possíveis, que vão exigir, no entanto, a admissão de novas entidades na sua ontologia. Uma destas alternativas consiste na postulação de *universais*, entidades totalmente presentes em várias localizações espaciotemporais. Os universais são úteis numa teoria das propriedades, pelo facto de

podermos explicar a naturalidade de uma propriedade através da presença do mesmo universal em todas as suas instâncias – o universal estaria localizado onde a propriedade fosse instanciada, como uma parte comum, não espacial e não temporal, dos vários objetos que a instanciam. A outra alternativa substituiria os universais por *tropos*, idênticos aos universais à exceção do não serem repetíveis, não poderem estar localizados totalmente em diversos pontos espaciotemporais. Assim, em vez de dizermos que o mesmo tropo está localizado nas várias instâncias de uma propriedade natural, dizemos antes que os vários tropos que aí ocorrem são duplicados (Lewis 1983b: 344-47, 1986b: 63-5).

A teoria que introduz os tropos deixa ainda como primitiva uma relação de semelhança: e relação de duplicação entre tropos, tal como deixa a teoria nominalista que não pretende admitir como primitiva a naturalidade de certas propriedades. No entanto, como D. Lewis nota em (Lewis 1986b: 65-6), a relação de semelhança entre tropos é menos problemática do que aquela que é estabelecida entre indivíduos. Os tropos não são idênticos em certos aspetos e diferentes noutros: ou são semelhantes ou não o são. Em vez da relação de semelhança complexa R acima descrita, podemos ter apenas uma relação diádica S. Define-se uma propriedade natural, a partir de S, como a classe de objetos nos quais estão localizados tropos tais que, entre quaisquer dois deles é verdadeira a relação S. Repare-se que S é transitiva, o que não poderia acontecer com uma relação análoga a S que se estabelece entre indivíduos.

Apesar de apresentar estas alternativas, D. Lewis não chega a optar por nenhuma delas.

Anexo 4

Em "General Semantics" (Lewis 1970b), é apresentada e desenvolvida a forma geral de um género de gramáticas que, D. Lewis acredita, podem ser usadas para especificar qualquer linguagem – natural ou artificial – que seja, claro, finitamente especificável. Uma gramática deste género tem um léxico e uma componente transformacional. O léxico gera estruturas complexas nas quais certas expressões são associadas a *significados*; enquanto que a componente transformacional vai derivar diferentes estruturas, de acordo com as regras especificadas para a gramática, a partir daquelas que são determinadas pelo léxico. São as estruturas geradas pela componente transformacional que vão associar as expressões da linguagem falada ou escrita aos significados determinados pelo léxico. Através destas componentes, diz-se que uma gramática especifica uma *relação de representação* entre os significados e marcas físicas que servem como expressões (Lewis 1970b: 35). Aqui, vou concentrar-me apenas na componente lexical.

O léxico é um conjunto finito de triplos ordenados (*e*, *c*, *v*) de uma expressão *e*, uma categoria *c* e um valor semântico *v* apropriado à categoria *c*. Apesar de estarmos a tratar de um conjunto finito, o seu propósito é o de determinar um número infinito destes triplos, através das operações sintáticas e semânticas associadas às categorias e aos valores das expressões.

De entre as categorias que ocorrem no léxico, distinguimos as categorias *básicas*, a partir das quais se obtêm as categorias *derivadas*, cujo número é infinito (apesar de apenas um número finito ocorrer no léxico, obviamente). Se c, c_1, …, c_n ($n > 0$) forem categorias – básicas ou derivadas –, (c / c_1, …, c_n) é uma categoria derivada. A partir daqui, é possível formular várias regras para a formação de expressões a partir das que ocorrem no léxico, da seguinte forma:

(R1) È admitida qualquer expressão que resulta da concatenação de expressões das categorias (c / c_1, …, c_n), c_1, …, c_n, nesta ordem, e a expressão resultante vai pertencer à categoria c.

Se todas as expressões formadas tiverem de se conformar a esta regra, podemos ver que o léxico tem de atribuir expressões a categorias derivadas, sob pena de não poderem gerar novas expressões (Lewis 1970b: 20-2).

Através deste sistema de categorias, foi possível especificar parte da componente sintática da gramática – o restante trabalho na sintaxe será desenvolvido pela componente transformacional. Resta, ainda, olhar para os elementos semânticos do léxico. Para fazer as relações semânticas entre as expressões espelharem as suas relações sintáticas, trata-se o valor semântico apropriado a uma categoria derivada (c / c_1, …, c_n) como uma função que liga uma sequência de n valores semânticos – correspondendo o primeiro a uma expressão da categoria c_1, …, e o n-ésimo a uma expressão da categoria c_n – a um valor semântico apropriado à categoria c (Lewis 1970b: 22-31). Temos, então, a seguinte regra semântica:

(R2) Se α pertence à categoria (c / c_1, …, c_n) e o seu valor semântico é v, β pertence à categoria c_1 e tem como valor semântico v_1, …, e η pertence à categoria c_n e o seu valor semântico é v_n, a expressão resultante de $\alpha + \beta + \ldots + \eta$ terá como valor semântico $v(v_1, \ldots, v_n)$.

A partir das considerações anteriores, estamos aptos a ver que triplos de um léxico que tenham a seguinte forma:

(e, (c / c_1, …, c_n), ϕ),
(e_1, c_1, v_1),
…,
(e_n, c_n, v_n).

Podem gerar um triplo como este:

$((e + e_1 + \ldots + e_n), c, \phi(v_1, \ldots, v_n))$.[88]

O sistema de categorias proposto em (Lewis 1970b: 20-2) é formado pelas categorias básicas dos nomes (N), dos nomes comuns (C), e, claro, a das frases (F). (Convém notar que o propósito de uma gramática é a de gerar frases e valores semânticos para as mesmas.) A partir destas podemos formar, por exemplo as seguintes categorias derivadas:

(1) Predicados: (F/N),

(2) Adjetivos: (C/C),

(3) Advérbios: ((F/N)/(F/N)),

….

Os valores semânticos apropriados para as categorias básicas são intensões – funções de índices para extensões. (Ou, melhor dizendo, uma função de contextos e índices para extensões. Esta alteração foi introduzida depois de (Lewis 1970b), em (Lewis 1980a), e é explicada na secção 1.5.4.) A extensão de uma frase é um valor de verdade, de um nome é o objeto por ele referido em várias circunstâncias, e de um nome comum é o conjunto de coisas a que este se aplica. Assim, a cada frase é associada uma função de índices para valores de verdade, a cada nome uma função de índices para objetos e a cada nome comum uma função de índices para conjuntos de objetos (Lewis 1970b: 22-7).

88 A simplicidade das regras de sintaxe e semântica aqui consideradas é, apesar de tudo, um entrave para o sucesso de uma gramática deste género, como é notado em (Lewis 1970b: 22). Por exemplo, é impossível acomodar a ordem das palavras apresentada nas frases da linguagem natural, e são admitidas inúmeras construções indesejadas. Estes problemas são resolvidos, no entanto, pela componente transformacional.