Intencionalidade: Mecanismo e Interacção

Porfírio Silva
Universidade Técnica de Lisboa

Abstract. In this essay we try an answer to the question *has intentionality to be reduced to anything?* We propose that it is possible to reduce any variety of intentionality to a specification of mechanisms (internal organization of the items involved in a given intentional phenomenon) and a historical pattern of interaction (structure of mutual significant relations historically acquired by different items involved in the same intentional phenomenon). We first clarify the meaning of this proposal having recourse to the Ruth Millikan's teleosemantics. Then, we assess the relevance and feasibility of our proposal, considering, in a succession, the case of animals and humans, of machines, and of sophisticated human collectives. We conclude arguing the heuristic nature of the proposed reduction.

Keywords: Intentionality, mechanism, interaction, reductionism, Ruth G. Millikan.

Muito trabalho filosófico respondeu durante mais de um século ao fogo ateado por Brentano (1874) quando sugeriu que a intencionalidade — a possibilidade de alguma coisa representar estados de coisas e visar objectos diferentes dessa mesma coisa, ainda que o estatuto ontológico do objecto intencional faça dessa relação uma quase-relação — é uma marca do mental: todos os fenómenos mentais, e só os fenómenos mentais, são capazes de intencionalidade. Não cabe aqui percorrer nenhum dos caminhos (nem tentar resolver nenhum dos becos) que essa tradição plural criou. O presente ensaio é dedicado exclusivamente a tentar uma resposta à seguinte questão: *tem a intencionalidade de poder ser reduzida a alguma coisa?*

O que se propõe desde já é que *é possível reduzir qualquer variedade de intencionalidade a uma especificação de mecanismos (organização interna dos itens intervenientes num fenómeno intencional) e um esquema histórico de interacção (estrutura das relações mútuas significativas adquiridas historicamente pelos vários itens intervenientes no mesmo processo intencional)*.

O que se fará seguidamente (secções 1 a 3) é tentar evidenciar o interesse e a viabilidade desta proposta, considerando, primeiro, o caso do mundo animal não humano e o caso dos humanos; depois, o caso das máquinas; e, finalmente, o caso dos colectivos sofisticados especificamente humanos. Só depois desse percurso, a terminar (secção 4), será considerada directamente a questão da necessidade dessa redução — momento em que se precisará que tipo de redução é esta que se propõe.

Esta proposta de redução da intencionalidade à equação "mecanismos mais interacção histórica" tomará por modelo o núcleo central da abordagem teleosemântica

Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

de Ruth Millikan, que nos parece ser a que ainda hoje melhor faz justiça a uma compreensão adequada do mundo natural a que pertencemos. Contudo, não é a abordagem inspiradora de Millikan que será aqui defendida, nem sequer muito discutida — mas sim, especificamente, a proposta de redução aqui formulada.

A história mostra que alguma da melhor filosofia sempre alimentou e se alimentou de outras formas de aproximação ao mundo (como a ciência e a política) — e é essa procura de confluência que aconselha esta redução da intencionalidade. Por isso, finalmente, ela será qualificada como heurística.

1. Intencionalidade para animais, os humanos e os outros

A proposta de Ruth Millikan é uma das que não se distraem do facto de os humanos serem criaturas naturais num mundo natural – por isso procurando compreender a intencionalidade como um fenómeno natural. Neste caso, a nossa moldagem pela evolução natural é encarada seriamente. Os principais textos da exposição inicial desta concepção (Millikan 1984; 1993) explicam a característica central dos signos intencionais — poderem significar algo não existente — com base num alargamento da noção de função biológica: a teoria das funções próprias, desenhada com abrangência suficiente para abarcar tanto as funções dos dispositivos biológicos como as dos dispositivos linguísticos. Vejamos.

A teleosemântica de Millikan deriva da concepção teleológica das funções biológicas: um dispositivo tem uma *função própria directa* se o sucesso (proliferação) da sua linhagem se deve em parte ao facto de, historicamente, essa família de dispositivos ter desempenhado essa função mais frequentemente do que certos outros dispositivos, por possuir certa característica com uma correlação positiva com o desempenho dessa função.

Ter uma teleofunção não é ter uma função no sentido do funcionalismo clássico. Os dispositivos só têm funções próprias enquanto membros de *famílias estabelecidas reprodutivamente*, e não devido à forma, ou às disposições, ou ao desempenho efectivo ou possível de um espécime num dado momento: um coração que não bombeia o sangue por ter uma malformação ou estar doente, não deixa por isso de ser um coração; poucos espermatozóides efectivamente realizam a sua função própria, mas têm-na. "Ter uma função própria depende da *história* do dispositivo que a tem, não da forma ou das suas disposições" (Millikan 1984: 29). E, por isso, um coração artificial não passa a ser um membro da categoria biológica "coração" por ter a forma de coração ou bombear sangue — porque tem a história errada. Os membros de uma família podem ser reproduzidos por outros membros da mesma família (espécimes do mesmo gene), ou por membros de outra família que tem como função própria essa reprodução (o meu coração não foi reproduzido pelos corações dos meus antepassados, mas por uma família de genes).

Quase invariavelmente, um dispositivo só desempenha as suas funções próprias num ambiente apropriado, frequentemente incluindo membros Normais de outras famílias reprodutivas (o coração e o resto do sistema circulatório só funcionam adequadamente em ligação) em certas condições (o olho só vê com certa luz). As *condições Normais* para o desempenho de uma função própria de um dispositivo são as condições em que historicamente estiveram os membros da respectiva família reprodutiva quando efectivamente desempenharam essa função. Mais genericamente, uma *explicação Normal* é esse tipo de explicação histórica, incluindo como é que certas características do dispositivo se relacionam com o desempenho das suas funções. Note-se que Millikan usa *Normal* ou *Normalmente* para indicar um carácter quase-normativo, com o sentido biológico ou médico, não o sentido estatístico.

Cabe sublinhar que Millikan usa "categorias biológicas" numa acepção mais lata do que o biológico, incluindo todas as categorias de funções próprias com explicação histórica da sua proliferação, abrangendo, nomeadamente, artefactos cujo design não é original mas reproduzido de outros para que sirvam as mesmas funções que os seus modelos; comportamentos herdados, ou que resultam de treino, ou de aprendizagem por tentativa e erro com recompensas positivas.

Um exemplo frequentemente explorado por Millikan — a dança das abelhas — ilustra convenientemente esta abordagem.

É importante para o sucesso da colmeia que as melhores fontes de alimento sejam exploradas intensamente e enquanto estejam nas melhores condições, em vez de cada obreira simplesmente explorar qualquer fonte que descubra. A "dança das abelhas" é instrumental no recrutamento de mais obreiras para os melhores recursos que alguma tenha identificado. Uma obreira, ao voltar para a colmeia com pólen ou néctar suficientemente nutritivo para justificar exploração intensa da respectiva fonte, executa perante as outras uma dança que mapeia o local da colheita. A dança mais elaborada tem a forma do algarismo oito: a abelha voa em linha recta numa curta distância, regressa em semicírculo ao ponto de partida, voa de novo o mesmo percurso em linha recta, depois faz outro semicírculo na direcção oposta ao anterior e completa um voo em forma de oito. Esta forma básica é executada repetidamente.

A duração da parte central da dança, em linha recta, é proporcional à distância entre a colmeia e a fonte de alimento. Outros elementos relacionam-se com a direcção. A orientação da linha central do "oito", o voo em linha recta, indica a direcção em relação ao sol: a linha desse voo tem, em relação à vertical do lugar, um certo ângulo; é o mesmo ângulo formado pela linha de voo que leva da colmeia ao campo de flores e pela linha que liga a colmeia à vertical do sol no momento. A dança transpõe um ângulo solar para um ângulo gravitacional. Por exemplo, se para ir da colmeia à fonte de comida as abelhas naquele momento tiverem de voar exactamente em direcção ao sol, a abelha executa uma dança em que o segmento central do "oito", em linha recta, estará na vertical e será executada de baixo para cima. Se o caminho

para as flores for exactamente o oposto, o voo será igualmente na vertical, mas de cima para baixo (Visscher 2003).

A intencionalidade dos elementos linguísticos consiste basicamente em eles traçarem mapas do mundo não linguístico: mas esse mapear existe claramente em outros domínios do mundo natural. É o caso da dança das abelhas, que exemplifica uma das variedades de intencionalidade, os ícones intencionais, ao satisfazer as características seguintes.

Primeiro, uma dança das abelhas é um membro de uma família reprodutiva (uma família de coreografias com uma sintaxe global invariante, que acomoda as variações responsáveis pelo conteúdo: as localizações de néctar), que tem uma função própria directa: recrutar abelhas para explorar o néctar.

Segundo, Normalmente uma dança das abelhas está entre dois dispositivos cooperantes, os mecanismos produtores da coreografia na abelha que dança e os mecanismos consumidores (interpretadores) nas que observam, mecanismos padronizados para ajustamento mútuo, sendo que a presença e cooperação de cada um faz parte das condições Normais para o desempenho próprio do outro.

Terceiro, Normalmente, um ícone intencional serve para a adaptação do dispositivo intérprete a condições que lhe permitam o desempenho das suas funções próprias. Na dança das abelhas, o mecanismo de interpretação das que observam permite-lhes a adaptação à localização da fonte de néctar em cada caso, produzindo um voo com direcção adaptada.

Quarto, Normalmente a dança das abelhas é um ícone intencional indicativo ao mapear alguma coisa no mundo (uma configuração real envolvendo néctar, Sol e colmeia) e um ícone intencional imperativo ao levar o dispositivo interpretador a produzir algo visado pelo ícone (abelhas recrutadas para colher néctar na localização visada).

As frases indicativas ou imperativas das linguagens humanas, que Millikan considera o paradigma central dos signos, também são ícones intencionais, satisfazendo igualmente aquelas características das danças das abelhas. São, contudo, também representações: são ícones intencionais que, quando desempenham as suas funções próprias, têm os referentes dos seus elementos identificados pelos intérpretes. É que nós sabemos acerca do que são as danças das abelhas, mas as abelhas não sabem: limitam-se a reagir a elas apropriadamente. Já uma frase, se também pode não ser compreendida pelos que a usam, só desempenhará a sua função Normalmente se o for. A variedade da intencionalidade também é concebida por Millikan noutra direcção: nem todas as frases são representações, mesmo quando satisfazem as suas funções próprias.

O exemplo da dança das abelhas permite clarificar outros paralelismos entre intencionalidade humana e animal.

Algumas das funções próprias directas de um dispositivo podem ser *funções pró-*

prias relacionais. O que importa numa dança das abelhas são relações: com uma configuração colmeia-Sol-néctar, com um voo a ser empreendido. Um dispositivo com uma função relacional adapta-se àquilo com que se relaciona (o seu adaptador) e, para isso, pode produzir outros dispositivos: *dispositivos adaptados*. Os dispositivos adaptados adquirem assim funções próprias derivadas. (Note-se que podem formar-se camadas e redes intrincadas de funções derivadas, permitindo a emergência de dinâmicas que se afastam muito da selecção genética directa.) Se uma abelha descobre néctar 700 metros a nordeste da colmeia, Normalmente tem mecanismos com a função própria derivada adaptada de coreografar a dança para essa localização específica. Mas se uma abelha faz uma dança que indica uma certa localização de uma fonte de néctar que não existe no mundo, essa dança é um *dispositivo inadaptado*. Dispositivos intencionais linguísticos podem revelar o mesmo tipo de inadaptação, levando Millikan a escrever que uma frase falsa não é mais problemática do que um espécime da dança das abelhas que mapeia uma fonte de néctar que não existe (Millikan 1984: 88)[1].

Nem sempre é linear individualizar um conteúdo intencional. Suponhamos que uma abelha descobre néctar a sudoeste da colmeia, mas, devido a uma anomalia no seu mecanismo de coreografia, produz uma dança que mapeia néctar a nordeste. O que representa essa dança? Para os mecanismos produtores, essa dança tem a função própria de recrutar abelhas para sudoeste. Mas, pelo lado dos mecanismos consumidores, as abelhas que observam serão Normalmente recrutadas para nordeste. O que prevalece é o que está historicamente na base da proliferação do dispositivo em causa: a função própria directa da dança enquanto membro da família das danças das abelhas sintacticamente correctas. E isso dá uma primazia ao mecanismo consumidor, numa relação Normal com a história evolutiva. O que aquela dança representa é esclarecido pelo facto de que, se outras abelhas a observarem, serão recrutadas para uma localização a nordeste. Esta abordagem ajudará a iluminar o funcionamento da linguagem humana como linguagem pública.

As abelhas precisam de viver onde haja flores, e vice-versa, pelo que o funcionamento Normal de certos dispositivos tanto das abelhas como das flores deve estar cruzado de forma a preservar essa ligação mútua. Esse cruzamento é servido por funções próprias de padronização e estabilização. O mesmo acontece na linguagem. As funções de estabilização são funções próprias directas de famílias de dispositivos linguísticos, ajudando a explicar porque é que os falantes continuam a falar usando desse modo esses dispositivos e os ouvintes continuam a reagir-lhes padronizadamente. Se os falantes não conseguissem dos ouvintes conformidade com as suas elocuções de frases imperativas, ou não conseguissem despertar crenças verdadeiras com as elocuções de frases indicativas, e isso não acontecesse de forma massiva em proporção das elocuções totais, esses tipos de elocução teriam desaparecido. De sublinhar aqui o facto de as funções de estabilização assentarem muitas vezes na

sintaxe, que fornece a entrada para a interpretação enquanto indicador da família do elemento linguístico em ocorrência.

Compreende-se assim que, numa linguagem pública, o que um dispositivo significa não é o que quer dizer o falante responsável pela elocução de um espécime num dado momento: esse falante pode não o compreender bem ou estar a dar-lhe um uso parasita. Se um dispositivo linguístico não tivesse as suas funções próprias, não poderíamos distinguir o que ele significa daquilo que um falante particular pretende significar com ele. Nem poderíamos dizer se o que alguém afirma é verdadeiro ou falso, se as funções das suas palavras dependessem das suas intenções (ou intensões).

As frases têm intencionalidade própria, fundada em relações naturais externas, resultado de relações Normais que essas frases suportam entre produtores e consumidores desses dispositivos. Os espécimes de famílias de dispositivos linguísticos podem ter funções próprias derivadas das intenções dos falantes, e por vezes elas não coincidem com as funções próprias directas desses dispositivos linguísticos, mas não são as intenções do falante que prevalecem. Prevalecem as funções de estabilização, porque elas é que explicam como esse dispositivo linguístico proliferou e chegou até esse falante.

Deve ser agora claro por que partimos de Millikan para pensarmos se a intencionalidade pode ser reduzida a *mecanismo* e *interacção histórica*. É que a abordagem desta autora assenta essencialmente na explicação pela história evolutiva dos dispositivos intencionais, sendo que a evolução é essencialmente interacção a produzir os seus efeitos no tempo, cumulativamente, num constante desenho de trajectórias com consequências. O que é trabalhado por essa história são mecanismos, moldando e moldados por essa interacção: os pêlos sensores das abelhas medindo a posição relativamente à gravidade, os olhos captando os padrões de polarização da luz no céu (dispensando a visualização do Sol) ou medindo distâncias em termos de fluxo óptico.

Nesta proposta radicalmente histórica, a interacção não pode sequer ser manufacturada. Podemos encaixar uma tese característica de Millikan no "argumento do Homem do Pântano": se cair um raio sobre um tronco podre num pântano e daí surgir instantaneamente um homem que seja, digamos, uma cópia física exacta do leitor – essa criatura não teria crenças, nem desejos, nem intenções, . . . , porque teria uma história evolutiva errada. Aliás, tão-pouco teria coração, fígado, olhos, cérebro — nenhuma função biológica, porque todas as categorias de funções próprias dependem da história evolutiva e não da presente constituição ou disposições, sendo insusceptíveis de criação (ou análise) instantânea.

O que significará isto fora do mundo biológico?

2. Intencionalidade para máquinas

Desde 1956 que a Inteligência Artificial (IA) é o espaço da mais sistemática tentativa para construir máquinas com algum tipo de mente — e, designadamente, máquinas com intencionalidade. Um dos problemas teóricos mais persistentes nessa linha de investigação consiste precisamente em saber se as máquinas, tendo alguma forma de intencionalidade, podem ter intencionalidade própria ou apenas intencionalidade derivada (por atribuição de intérpretes humanos).

O paradigma central do programa de investigação da IA clássica é a "hipótese do sistema simbólico físico" (HSSF), cuja formulação canónica se deve a Allen Newell e Herbert Simon (Newell e Simon 1976; Newell 1980), para quem a HSSF constitui solução para "aquilo a que os filósofos chamam o problema da intencionalidade": "como é que os símbolos num sistema simbólico representam algo externo ao sistema simbólico". Os banais computadores digitais electrónicos constituem o exemplo mais familiar dos sistemas simbólicos físicos (SSF) de que a HSSF é uma teoria — mas a HSSF abrange algo mais. Vejamos.

Sendo os símbolos conjuntos de padrões físicos susceptíveis de certas relações físicas entre si (permitindo combinar espécimes em expressões), um SSF é uma máquina que, por aplicação sucessiva de processos modificativos, produz no tempo séries de estruturas simbólicas. Num exemplo do que seria um SSF, estes autores colocam uma memória, que armazena um conjunto de expressões que constituem as referências de um conjunto de símbolos; um conjunto de operadores que processam símbolos; um controlo que aplica um operador à expressão simbólica activa; uma via receptora para novas expressões que descrevem o ambiente externo; certas ligações entre operadores e órgãos motores produtores de comportamento externo (Newell 1980: 142–7).

Então, segundo a *HSSF, um SSF tem os meios necessários e suficientes para a acção inteligente geral* (Newell e Simon 1976: 116). A HSSF associa-se explicitamente à ideia de que quer humanos quer computadores são instâncias de SSF e que os símbolos dos computadores e os dos humanos são os mesmos (Newell 1980: 135–6).

Como um SSF é uma máquina que existe num mundo de objectos mais vasto do que o conjunto das expressões simbólicas, precisamos de duas noções centrais para compreender a relação de intencionalidade entre símbolos e outros objectos: *designação* e *interpretação*. Ora, na explicação do mencionado exemplo, embora se fale de órgãos "receptores" e "motores", parecendo haver ligações de e para o mundo, essa ilusão desfaz-se quando se explicitam as noções de "designação" e "interpretação". A *designação* ("Uma entidade X designa uma entidade Y relativamente a um processo P se, quando P toma X como input, o seu comportamento depende de Y"), supostamente, dota o sistema de uma "acção a distância" (na expressão de Newell), porque o comportamento do sistema não é uma função dos símbolos propriamente ditos,

mas uma função das entidades que os símbolos designam. Só que, quando vamos à descrição técnica do operador que implementa a designação ("acesso"), percebemos que as relações de acesso que podem ser criadas são apenas as que ligam símbolos dentro da máquina a outras entidades dentro da mesma máquina (Newell 1980: 156, 160). Quando se trata de descrever a *interpretação* — "o acto de aceitar como input uma expressão que designa um processo e então executar esse processo" (Newell 1980:158) — afirma-se que os símbolos que designam operadores são essenciais, porque contêm uma semântica externa, apontando para comportamentos que incorporam o sentido que as operações do sistema fazem no mundo exterior. Só que não é dada qualquer explicação acerca de como isso se produz. Porque não se produz — e assim continuará nas sucessivas reelaborações desta proposta.

Fodor foi cortante nesta questão. Comentando o robot simulado SHRDLU, para quem Winograd tinha programado um "mundo" acerca do qual o robot pudesse "falar", critica a pretensão de que as frases das linguagens de programação ganhem uma semântica genuína quando interpretadas para linguagem-máquina, por esta ligar directamente "ao mundo", isto é, aos estados físicos mais elementares do computador. Interpretar desse modo, digamos, a frase "Boise é uma cidade" é dizer que a expressão "BOISE" aponta para um endereço de memória com o rótulo "CIDADE". Isso não é mais do que pretender que "Napoleão venceu a batalha de Waterloo?" quer dizer "Verifique se a frase 'Napoleão venceu a batalha de Waterloo' ocorre no volume que tem o número XXX,XXX na numeração decimal de Dewey na Secção da Rua 42 da Biblioteca da Cidade de Nova York" (Fodor 1978: 204–11).

Haugeland (1985) identifica o núcleo duro da IA clássica com a HSSF, que traduz na tese de que tanto os computadores como os humanos são sistemas formais automáticos interpretados — mas precisa que essa tese depende essencialmente de outra, a "divisa formalista". Tendo os espécimes de símbolos num sistema formal "duas vidas" — uma "vida sintáctica", na qual são marcas sem significado, manipuladas exclusivamente de acordo com as regras internas do jogo, e "uma vida semântica", na qual têm significados apontando para o mundo exterior — a "divisa formalista" é: "Trata da sintaxe, que a semântica trata dela própria" (Haugeland 1985: 106). Quer dizer: aceites como verdadeiros os axiomas do sistema formal, se as regras de inferência preservam a verdade, então qualquer processamento pelo "sistema formal automático interpretado" de uma fórmula com sentido à entrada resultará, à saída, numa fórmula com sentido na mesma acepção. Quando uso uma calculadora, o resultado obtido carregando na tecla "=" tem sentido debaixo da mesma interpretação que empreguei para escolher o arranjo de teclas com que inseri os dados, ordenadas pela pergunta que pedia aquela resposta. O problema desta leitura é que renuncia directamente a qualquer forma de intencionalidade para máquinas que não seja meramente derivada.

Entretanto, a partir de Millikan conseguimos ver o impacte da falta de uma his-

tória de interacção. A "divisa formalista" ignora radicalmente a natureza das funções próprias da sintaxe, as "biológicas" funções de estabilização e padronização essenciais a uma linguagem pública, que emergem precisamente da interacção. Porque a "vida da sintaxe" não é só mecanismo, é também interacção moldada na e pela história evolutiva.

A IA começou a enfrentar resolutamente este problema depois da sua formulação explícita por Stevan Harnad (1990), que o baptizou como "problema da fundação dos símbolos": como é que a semântica de um sistema formal automático interpretado podia ser intrínseca e não parasita dos intérpretes humanos, se um computador digital com programa armazenado está face ao mundo como alguém tentando aprender chinês como primeira língua a partir do zero, apenas usando um dicionário chinês-chinês, ou mesmo todas as obras existentes escritas em chinês, mas nada além disso: nem outras linguagens, nem qualquer experiência acerca do mundo (Harnad 1989). A sua resposta implicava que um SSF carecia de algum subsistema capaz de captação sensorial, pelo qual o mundo exterior impressionasse por via não simbólica o processamento simbólico.

Posteriormente, fará uma revisão importante da sua formulação do problema, diagnosticando-lhe um enviesamento internalista: atribuindo aos símbolos nas máquinas uma semântica derivada dos intérpretes humanos, implicitamente coloca a semântica dos humanos "dentro da sua cabeça". Em (Harnad 2002) explica que o computacionalismo, após ter sofrido de uma teoria do significado baseada no conteúdo mental estrito, em que só conta o que está "dentro da cabeça" (da pessoa ou da máquina), ao compreender o interesse de considerar o conteúdo lato (a parte do mundo exterior no significado) cometeu outro erro: tentar escrever um modelo do mundo e programá-lo directamente na máquina — algo como escrever o modelo do mundo "dentro da cabeça da máquina", o que se revelou igualmente improdutivo.

Claramente, mesmo os críticos da IA clássica insistiam no erro de considerar apenas o mecanismo, tardando em descobrir que precisavam de contar também com a interacção. Andavam a querer construir máquinas intencionais com os mesmos problemas do "Homem do Pântano": instantâneas, sem história de interacção. Alguns tentavam esboçar a alternativa interaccionista: "Proponho pensar acerca da computação em termos de maquinaria e de dinâmica. Uma máquina (...) é um objecto no mundo físico que obedece às leis da física. (...) [A dinâmica] diz respeito às interacções entre um indivíduo (robot, formiga, gato ou pessoa) e o seu ambiente circundante" (Agre 1997: 53, 57–9). Contudo, o elemento histórico continuava mal compreendido. Recente alento para esta via veio da Nova Robótica.

Um dos pressupostos filosóficos mais importantes da IA clássica é o funcionalismo, caução metodológica para algumas das opções mais desastrosas desse programa de investigação, particularmente no que toca ao desprezo generalizado pela questão da realização física da mente das máquinas. Um texto essencial na implan-

tação do funcionalismo na caixa de ferramentas intelectuais da IA, (Putnam 1960), surge a tempo de alcançar uma proeminência perniciosa que só a custo, décadas volvidas, começará a ser ultrapassada pela "nova IA", com robots físicos em ambientes físicos. Nessa "Nova IA" (Robótica) jogou inicialmente papel destacado Brooks (1999), que tenta dispensar os símbolos, concentrar-se nos comportamentos, dar aos sistemas perceptivo e motor o trabalho da "fundação física" dessas "criaturas", tentando assim uma ligação tão directa ao mundo exterior que a própria necessidade de representações é suprimida ("o mundo é o seu melhor modelo"). Felizmente, a nova IA Robótica sobreviveu a essa estratégia radical, que se revelou igualmente incapaz de compreender uma inteligência pelo menos tão sofisticada como a humana (Steels 2003). Interessante aqui é que a compreensão biológica das funções podia ter evitado alguns grandes problemas à IA: é que as funções, não sendo causas, mas efeitos — efeitos de uma história evolutiva — não podem ser instaladas "à mão" por projectistas humanos, instantaneamente, em máquinas que assim ficam, na expressão de Millikan, com "a história errada" — e, assim, com mecanismos incapazes de intencionalidade.

Não obstante, a maioria dos robots continuam a padecer da mesma doença: saem directamente da mão do artesão, ou da linha de montagem, e não têm qualquer história evolutiva. Pode esperar-se, contudo, uma futura viragem millikaniana na robótica, por via da Robótica Evolucionista (Nolfi e Floreano 2000).

A Robótica Evolucionista procura alternativas a que os sistemas robóticos sejam directamente projectados por humanos, recorrendo a ferramentas (como o algoritmo genético) inspiradas na evolução natural para desencadear processos de evolução artificial que sejam os responsáveis pelo desenho de aspectos importantes de um robot (do seu "cérebro" ou do seu "corpo"). A ideia é criar possibilidades de que os enviesamentos que o projectista humano impõe (explícita ou implicitamente) na construção propositada de uma série de robots, sejam substituídos pelos enviesamentos induzidos pelas características da plataforma, do ambiente e da tarefa imposta à "criatura" — criando mesmo surpresa de origem evolutiva (artificial) aos responsáveis pela definição dos parâmetros de partida.

Sem tentar aqui um balanço dessa linha de investigação recente, sempre cabe enunciar uma precaução. Essas máquinas, mesmo que venham a ter intencionalidade própria, o que não excluímos, não chegarão a ter a mesma linguagem pública que nós sem que a sua linhagem evolutiva entre em relação com a nossa própria linhagem evolutiva — se não entrarmos no mesmo esquema de interacção como dispositivos cooperantes em que elementos dessa linguagem fazem parte do nosso ambiente Normal. Se Millikan tiver razão. Como teve quando mostrou que não bastava o mecanismo, sem interacção. Como a IA clássica insistiu em não compreender.

3. A realidade institucional e a intencionalidade colectiva

Temos até aqui, inspirados em Millikan, sustentado a produtividade de reduzir a intencionalidade a mecanismo e interacção, quer no caso de animais e humanos, quer no caso de máquinas. Poderá esta proposta ser útil na compreensão da intencionalidade colectiva, especialmente no plano da realidade institucional, que parece especificamente humana?

Uma entrada incontornável para pensar estas questões é a obra de John Searle sobre a construção da realidade social, a começar por (Searle 1995). A pedra basilar desse exercício é a distinção entre factos brutos e factos institucionais. Um facto bruto é um facto cuja existência nada deve aos observadores (o pico do monte Everest está a N metros de altitude). Um facto social é um facto que envolve, antes da intencionalidade individual, intencionalidade colectiva: estou a tocar violino como parte da orquestra estar a tocar uma sinfonia; a orquestra a tocar a sinfonia não é um sucedâneo de uma colecção de executantes a tocar partes da peça. Um grupo de hienas a caçar um leão, o que não funcionaria sem implicar um grupo de forma coordenada, é outro exemplo. Os factos institucionais são um subconjunto dos factos sociais. A criação de factos institucionais envolve os mecanismos pelos quais um colectivo decide atribuir certa função a certo tipo de objectos, sendo que essa função não podia decorrer apenas das características físicas (ou químicas ou biológicas) desse objecto e tem de ser activada pela cooperação continuada entre os indivíduos desse colectivo.

Partindo desta distinção, (Searle 2006) sistematiza uma abordagem à realidade institucional como realidade especificamente humana, assente em três pilares.

Primeiro, a intencionalidade colectiva. Além da intencionalidade individual existe intencionalidade colectiva, descritível por formas como “Nós desejamos”, “Nós cremos”, “Nós tencionamos”. A intencionalidade colectiva pode apresentar-se, nomeadamente, como acção intencional colectiva (tocar violino como parte de tocar a sinfonia) ou como crença colectiva (uma comunidade religiosa recitando o Credo de Niceia expressa uma crença identificadora). Como vimos, Searle define os factos sociais como qualquer facto envolvendo intencionalidade colectiva de dois ou mais agentes humanos ou animais.

Segundo, as funções de estatuto. Os humanos, bem como certos animais, têm a capacidade para atribuir funções a objectos. Se uma pessoa pode usar um cepo como cadeira, um grupo pode usar um tronco como banco. Aqui, a atribuição funcional é suportada em características físicas dos objectos. Os humanos, parece que exclusivamente, são capazes de atribuições funcionais para as quais as características físicas do objecto são largamente irrelevantes. Nesse caso, falamos de funções de estatuto. O dinheiro, como função, não depende do suporte físico escolhido para notas ou moedas, apesar de certos critérios terem relevância prática (facilitar o transporte,

dificultar a falsificação). Aliás, o "dinheiro electrónico" consegue uma vasta desmaterialização do suporte, reduzido a algarismos em registos, sem prejuízo da função. Uma fronteira pode ter começado por ser assinalada por um muro, mas, enquanto condicionamento institucional, pode subsistir ao seu desaparecimento físico. A moeda e as fronteiras funcionam graças ao estatuto que lhe foi atribuído colectivamente pelos humanos. Em geral, factos institucionais e instituições são criadas por funções de estatuto atribuídas por actos de intencionalidade colectiva.

A forma geral de uma atribuição de função ("regra constitutiva") é "*X* conta como *Y* no contexto *C*". O dinheiro é uma instituição em que um certo tipo de pedaço de papel, produzido em certas circunstâncias, desempenha uma função que poderá ser descrita como "equivalente geral das trocas". O casamento é uma instituição em que certas palavras, proferidas pela pessoa certa nas circunstâncias previstas, valem como início de um certo tipo de relação entre as pessoas envolvidas. Em qualquer caso, as características físicas do termo *X* não criam o esperado estatuto: uma nota fisicamente idêntica às notas do banco central, mas produzidas à sua revelia, são falsas (como mostra o caso Alves dos Reis); as palavras "que casam", ditas por pessoa desprovida dos poderes apropriados, não casam. Duas propriedades formais das regras constitutivas são decisivas para compreender a sociedade humana: podem ser indefinidamente iteradas de forma ascendente (enquadramos instituições em novas instituições) e lateral (redes de instituições interligadas).

Terceiro, os poderes deônticos. O que é importante nas funções de estatuto é que elas são veículo de poder na sociedade. Aceitando funções de estatuto aceitamos um conjunto de normas que dizem respeito ao que é obrigatório ou permitido. Ficamos, desse modo, imersos numa rede de poderes deônticos. A propriedade, ou o casamento, dá-me específicos direitos e deveres. Este tipo de relações não existe no mundo animal, onde não há deontologia: um grupo animal pode ter um macho alfa, e isso implica certas relações de poder dentro do grupo, mas elas não resultam de funções de estatuto atribuídas colectiva e simbolicamente pelo grupo. Ora, obrigações e permissões são a fonte de razões para agir que não dependem de desejos: reconhecer que eu sou proprietário deste terreno dá às pessoas certas razões para agirem de certa maneira, razões essas não baseadas em desejos.

Ora, as formas especificamente humanas de socialização resultam desta combinação de funções de estatuto, poderes deônticos e razões para agir independentes de desejos (Searle 2006:19) — combinação assente na intencionalidade colectiva.

Mas é preciso ir um pouco atrás (Searle 1990) para compreender esta "intencionalidade colectiva". Aí, analisa-se o fracasso das abordagens sumativas à intencionalidade colectiva, enquanto tentativas de a reduzir à conjunção de intencionalidades individuais.

Searle assinala que o mesmo tipo de movimentos corporais pode formar, numa ocasião, um conjunto de acções individuais e, noutra ocasião, uma acção intencio-

nal colectiva. Exemplificando. Está um certo número de pessoas sentadas em vários pontos do relvado de um parque e, começando repentinamente a chover, todas se levantam e correm para o único abrigo disponível. Noutro momento, as mesmas pessoas fazem movimentos corporais indistinguíveis dos anteriores, mas executando a coreografia de uma peça de dança da companhia a que pertencem. Então, duas colecções indistinguíveis de movimentos corporais são, ora apenas comportamentos intencionais individuais, ora comportamento intencional colectivo, sendo este irredutível àqueles. Então, deve haver qualquer coisa de mental específico da acção colectiva — aí entra a intencionalidade colectiva.

É analisada também uma versão mais sofisticada da redução, desta vez a conjuntos de acções individuais suplementadas com conjuntos de crenças mútuas acerca das intenções dos outros membros do grupo — mas com o mesmo resultado negativo. Searle usa a noção de cooperação para resumir: a razão geral para a impossibilidade de reduzir intenções colectivas a intenções individuais é que a intencionalidade colectiva envolve a intenção de cooperar com outros membros de algum grupo.

Searle recusa uma concepção sumativa da intencionalidade colectiva — mas qual é a sua concepção? Searle compara a intencionalidade individual, da forma "Eu tenciono...", com a intencionalidade colectiva, da forma "Nós tencionamos...". Assim, podemos falar de "intencionalidade-eu" e de "intencionalidade-nós". Contudo, é preciso clarificar o que é a intencionalidade-nós. Para o individualismo metodológico de Searle, a sociedade consiste apenas em indivíduos, sem lugar para "mentes de grupo". Além disso, requer que a estrutura da intencionalidade de qualquer indivíduo seja independente da eventualidade de ele estar radicalmente enganado acerca do que realmente está a acontecer: em "nós tencionamos...", o "nós" pode não ter nenhum nós como referência. De acordo com este solipsismo metodológico, a intencionalidade-nós é intencionalidade exclusivamente de agentes individuais, eventualmente alucinados: pode ser intencionalidade de um cérebro numa cuba. Assim, tanto a intencionalidade-nós como a intencionalidade-eu são intencionalidades de indivíduos, embora diferentes: a intencionalidade colectiva, e só essa, contém a intenção de agir cooperativamente com outros. Como pode, então, Searle falar de intencionalidade colectiva ligada à cooperação? É que a intencionalidade colectiva assenta num fundamento biológico primitivo, também presente noutras espécies animais (dois pássaros juntos a construir um ninho), que identifica o outro como um candidato à cooperação.

Margaret Gilbert (2007: 40–5) apresenta uma das críticas mais interessantes à proposta searleana. O ponto de partida é a necessidade de distinguir as intenções-nós de Searle (só os indivíduos têm intenções-nós, os grupos não têm intenções) e as intenções-nós numa possível acepção alternativa (só os grupos poderão ter intenções de grupo). O que resulta de perguntarmos pela "intencionalidade-eles", correspondendo a expressões do tipo "Eles tencionam..."? Podemos, por exemplo, dizer "Eles

tencionam tocar a Quinta Sinfonia de Beethoven". Estaremos com isso a referir-nos apenas a uma colecção de intenções-nós dos membros da orquestra em causa, como Searle? Ou estaremos a referir-nos a uma intenção da orquestra como um grupo, coisa cuja existência Searle desmente?

Voltemos ao exemplo das pessoas no relvado e da dança. Searle pretende mostrar que num caso há apenas comportamentos intencionais individuais, enquanto no outro há comportamento intencional colectivo — e que, aqui, cada indivíduo tem a apropriada "intenção-nós", referindo a acção colectiva, e não apenas a intenção-eu, de fazer aqueles movimentos que constituem a sua parte na coreografia. Só que há mais possibilidades do que Searle vê. Se cada um daqueles mesmos indivíduos, sem pertencerem a qualquer companhia de dança, sem nunca terem combinado nada entre si, sem que exista qualquer coreografia, mesmo assim tiver as alucinações adequadas — se cada um alucinar que pertence a uma companhia de dança, que está a executar uma peça com tais e tais características, onde a sua parte é a mesma que seria na dança "real" do exemplo — e todos agirem de acordo com essas alucinações, qual será o resultado? Será a execução de um conjunto de movimentos corporais indistinguíveis, tanto da fuga desordenada da chuva como da apresentação de uma dança por uma companhia, movimentos corporais esses acompanhados das intenções-nós que existiriam no caso de haver intencionalidade colectiva — mas, realmente, não havendo grupo, nem acção de grupo, nem intencionalidade colectiva. Isto quer dizer que, assumindo os mesmos pressupostos que Searle, a intencionalidade colectiva não pode ser apenas uma colecção de intenções-nós (sempre intenções individuais).

A questão, para Gilbert, é que é duvidoso que possamos compreender o que é uma intenção-nós individual desligada de uma intenção de grupo. Eu não posso, apropriadamente, ter uma determinada intenção-nós sem ter a correspondente crença de que o mesmo "nós" tenha uma intenção colectiva com o mesmo objecto. Essa crença pode ser verdadeira ou falsa, ou alucinação — mas, se eu não tiver a crença de que a minha orquestra tenciona tocar a Quinta Sinfonia de Beethoven, não posso ter, como indivíduo, a intenção-nós expressa pela enunciação sincera de "Nós tencionamos tocar a Quinta Sinfonia de Beethoven".

Para Gilbert, há algo de fundamentalmente errado na abordagem solipsista à intencionalidade colectiva. As intenções de um grupo não podem ser fenómenos puramente mentais, não são algo apenas acerca de um conjunto de cérebros em cubas: têm de contar com algo que acontece no mundo exterior à mente. Para compreender a intencionalidade colectiva é preciso contar com compromissos conjuntos entre um certo número de agentes, o que não dispensa algum tipo de comunicação entre as partes.

Alguns trabalhos de Michael E. Bratman sobre o que chama "intencionalidade partilhada" ajudam a compreender o quão artificial é falar em intencionalidade co-

lectiva de cérebros numa cuba, que seria uma forma satisfatória de descrever a proposta de Searle. Para Bratman, a intencionalidade partilhada tem de incluir interacção e articulação entre intenções de uns e intenções de outros — e para isso têm de existir "outros". A concepção de intencionalidade que sugere toma-a como um elemento do planeamento da acção conjunta cooperativa. Bratman (1992) identifica três características cumulativamente necessárias à existência desse tipo de actividade: (i) prontidão mútua para agir em resposta à acção dos outros participantes; (ii) compromisso com a actividade conjunta; (iii) compromisso de apoio mútuo. Para que exista actividade cooperativa partilhada todas estas características têm de estar presentes e na ligação correcta: as atitudes mencionadas em (ii) e em (iii) têm de traduzir-se na prática prevista em (i), que só assim, centrada no compromisso que estabiliza uma rede de intenções, pode dizer-se cooperativa.

Esta abordagem envolve intenções que têm por objecto a actividade conjunta. Uma objecção seria que não posso intencionar as acções de outros. Mas posso. Não posso "tentar" as acções de outrém, nem tentar as "nossas" acções — mas posso intencioná-las. Essas intenções orientadas para o futuro têm um papel como elementos de articulação de planos parciais, para considerar os meios em ordem aos fins, para integrar os constrangimentos na reflexão prática — não fazendo sentido tentar reduzir as intenções ao papel de disparadores directos de acção imediata. Um sistema tal de entrelaçamento dinâmico de intenções, planos parciais e acções — visa o colectivo e é a forma adequada de conceber a intencionalidade colectiva.

Ora, esta análise retira plausibilidade ao aspecto internalista da concepção searleana da intencionalidade, que é um solipsismo metodológico. Parece-nos, contudo, que esta implausibilidade pode ser reparada quebrando a sua ligação a outro dos pressupostos metodológicos de Searle, o individualismo metodológico que o impede de considerar seriamente a possibilidade de "mentes de grupo". Vejamos como Pettit credibiliza essa possibilidade.

Philip Pettit, que rejeita interesse à alternativa dualista "individualismo ou colectivismo", julga podermos identificar "grupos com mente própria" (Pettit 2003). Vejamos como, partindo do "paradoxo doutrinal", originado no mundo jurídico.

Seja que um tribunal de três juízes deve decidir, num processo de indemnização por danos, pela responsabilidade do arguido se e somente se der por provado, cumulativamente, que uma omissão sua foi causa do dano ao queixoso e que o arguido tinha dever de assistência ao queixoso — tendo, estudado o caso, as seguintes opiniões:

	Causa do dano?	Dever de assistência?	Responsabilidade?
Juiz A	Sim	Não	Não
Juiz B	Não	Sim	Não
Juiz C	Sim	Sim	Sim

Com esta matriz, para conhecer a sentença temos de saber qual o procedimento de decisão adoptado. Num procedimento centrado na conclusão, o tribunal determina-se pelo voto dos juízes quanto à apreciação global do caso: aqui o arguido não será condenado. Num procedimento centrado nas premissas, primeiro é apurada a opinião do colectivo acerca dos pressupostos da sentença, daí resultando a sentença segundo as regras admitidas — neste caso resultando na condenação do arguido. O "paradoxo doutrinal" consiste em que, com as mesmas regras substantivas e com as mesmas opiniões de cada um dos juízes, seguir um ou outro dos procedimentos de decisão conduz a sentenças opostas.

Nesta situação o paradoxo produz-se sobre um caso em que a conclusão depende da conjunção das premissas, mas algo similar pode ocorrer dependendo a conclusão de uma disjunção. É solicitada a anulação de um processo em que o arguido tinha confessado e tinha sido considerado culpado. Para obter a anulação basta que as provas tenham sido obtidas ilegalmente ou que a confissão tenha envolvido coacção. Agora a situação é a seguinte:

	Provas ilegais?	Confissão forçada?	Anulação?
Juiz A	Sim	Não	Sim
Juiz B	Não	Sim	Sim
Juiz C	Não	Não	Não

Também nesta situação, um procedimento centrado nas conclusões conduzirá a um resultado (anulação) e um procedimento centrado nas premissas ao resultado oposto.

Generalizando: premissas apoiadas por maiorias não coincidentes podem obter a vitória num procedimento centrado nas premissas mesmo que não possam evitar a derrota num procedimento centrado nas conclusões, desde que a intersecção das maiorias não seja ela mesma maioritária.

Além desta generalização numa perspectiva sincrónica, Pettit considera outras formas de generalização, como o "dilema discursivo" para situações em que indivíduos pertencentes a um colectivo participam em séries de decisões desse colectivo ao longo do tempo, na presença de constrangimentos e do requisito de consistência da série de decisões colectivas. O que é essencial é que o "dilema discursivo" pode ocorrer em muitas situações de decisão colectiva lidando com questões racionalmente ligadas, de tal modo que possam formar-se sucessivas maiorias incoerentes. No limite, pode ser impossível a partir do momento t tomar qualquer decisão coerente com a série antecedente — mesmo que todas as decisões individuais tenham sido, enquanto tal, perfeitamente racionais. O que importa aqui é que a racionalidade da decisão colectiva não emerge espontaneamente da racionalidade da decisão individual.

Isto será particularmente sensível quando a incoerência comportamental do colectivo o impeça de atingir os seus objectivos colectivos (por exemplo, por descredibilização). Muitas vezes, defende Pettit, a única solução para esse risco consiste em assumir uma "razão colectiva", uma forma de decisão focada na preservação da coerência das diferentes decisões do colectivo, mesmo implicando alguma tensão entre razões individuais e colectivas.

Assim, pode tornar-se necessário falar propriamente de "grupos com mente", em colectivos como sujeitos intencionais distintos dos seus membros, numa disciplina da razão colectiva, no colectivo como interlocutor de outros agentes, em "pessoas institucionais" (Pettit 2003: 178–84).

Isto coloca a questão dos mecanismos que, ao nível colectivo, permitam gerir esta dinâmica. O dilema discursivo, que é apenas uma das dificuldades de agregação de juízos em séries de decisões sobre questões logicamente conexas (List 2006), exemplifica a tensão entre dois critérios: garantir um adequado nível de sensibilidade da posição do grupo às posições dos membros; garantir a coerência do grupo. O problema pode ser, por exemplo, que um elemento fundamental para admitir a razoabilidade de um sistema de decisão esteja em ele permitir a cada membro determinar-se autonomamente face a cada questão pela sua visão do mérito exclusivo dessa mesma questão — enquanto, para garantir a consistência das posições do grupo, essa garantia pode ser limitada com base no histórico das decisões colectivas. Isso coloca dois tipos de desafios. Primeiro, o grupo tem de tomar meta-decisões acerca dos próprios métodos de decisão do grupo. Segundo, o grupo, como agente intencional, tem de dotar-se de mecanismos de retroacção ao nível do colectivo (nomeadamente de monitorização do histórico de decisões) e não apenas ao nível dos indivíduos (Pettit 2007).

Chegados a este ponto temos de começar a tentar responder à questão com que abrimos esta secção. Será que a proposta de Millikan é capaz de dar conta da intencionalidade colectiva e da realidade institucional? Enquanto uma das tentativas de naturalização da intencionalidade, a sua ambição é fazer isso no quadro das ciências naturais — "a saber, a física, a fisiologia, a biologia, e a teoria da evolução" (Millikan 1984: 87). Brilha pela ausência nessa lista qualquer menção a uma ciência da sociedade. E isso tem consequências na proposta apresentada.

Apesar dos dispositivos intencionais serem incompreensíveis sem a sua vida pública (linguagem pública), os agentes considerados são sempre e apenas indivíduos. Qualquer eventual inclusão de "pessoas institucionais" ou "grupos com mente" como sujeitos intencionais careceria de uma reconsideração de toda a proposta. E talvez isso não possa ser excluído, se Pettit estiver certo. Os argumentos apresentados contra a irredutibilidade da intencionalidade colectiva à intencionalidade individual tornam plausível que a explicação desta não possa aplicar-se directamente à explicação daquela. Não há razão para assumir à partida que o enquadramento evolucionista

não seja capaz dessa adaptação, havendo muito trabalho teórico a atestar essa possibilidade. Mas essa articulação teria de fazer as suas provas. E teria de enfrentar problemas sérios em aberto, como o problema da acção estratégica, que tem resistido a um tratamento completamente satisfatório em termos individualistas, mesmo com ferramentas teóricas tão potentes como os modelos da Teoria dos Jogos. Por outro lado, sendo os factos institucionais criados por atribuição de estatutos largamente independentes da fisicalidade dos objectos, é aberto um horizonte ontológico para objectos intencionais que é intratável de forma directa pelos dispositivos que estão no foco da proposta de Millikan (ajustados como estão ao modelo dos dispositivos perceptivos). O papel dado por Searle às instituições na criação de razões para agir que não dependem de desejos constitui, por si só, um desafio directo às abordagens naturalistas, nas quais os desejos (e o correspondente modo imperativo dos signos intencionais) são o motor dos agentes. Afinal, tudo o que pudesse ser alinhado genuinamente num "ambiente institucional", num sistema de factos institucionais e instituições, poderia tornar menos nítida a leitura de propostas destinadas a lidar principalmente com "ambientes naturais". Como sublinha Searle, as instituições, além de constituírem um aspecto da realidade que está para lá da realidade física, servem também para nos libertar de certos constrangimentos físicos: a instituição "propriedade" permite que a posse de um bem não tenha que ser protegida à vista pela força; a instituição "casamento" permite que duas pessoas tenham uma relação de um certo tipo, originalmente ligada à habitação comum, mesmo que não estejam permanentemente em coabitação. E esse nível da realidade não encontra uma forma "natural" de explicação na proposta de Millikan.

Contudo, e por outro lado, também na consideração da realidade institucional e da intencionalidade colectiva, a tentativa de compreender a intencionalidade como "mecanismo mais interacção" se revela útil. No caso das tentativas da IA clássica para dotar as máquinas de intencionalidade, detectámos uma falta de atenção à componente interacção, a favor de uma exclusiva focagem no mecanismo — que se revelou contraproducente. Agora, no campo da realidade institucional, como forma especificamente humana de realidade social, podemos apontar os inconvenientes de considerar apenas a interacção (que vimos vários autores sublinharem como elemento da acção colectiva). O que se trata é de evitar um excessivo liberalismo na atribuição de mente e intencionalidade a colectivos. Um exemplo desse risco é a "consciência de classe", teorizada por exemplo por Gyorgy Lukács (1920), que mistura marxismo com influências hegelianas para conceber a existência de colectivos (classes sociais) integrados num colectivo mais vasto (sociedade) que podem, pela sua acção (luta política), contribuir para uma certa evolução histórica desse colectivo mais vasto (a revolução, o fim da sociedade de classes) desde que sejam capazes de se representar as condições objectivas da sua situação e as possibilidades objectivas que ela contém para a sua acção vitoriosa, representação essa que se dá numa espé-

cie de mente colectiva distinta da colecção das mentes individuais (consciência de classe). A este uso especulativo de noções para colectivos pode opor-se eficazmente a forma concreta como, por exemplo, Pettit especifica o que entende por "um grupo com a sua própria mente". Opor a pergunta pelo mecanismo a qualquer pretensão de ler um conjunto como um colectivo com genuína intencionalidade – pode ser um uso terapêutico da noção de mecanismo como elemento básico da intencionalidade. "Como se especifica o mecanismo que justifica essa forma de falar?" — pode ser uma questão clarificadora.

Então, o que estamos a sugerir é que, embora o naturalismo de Millikan possa enfrentar dificuldades específicas para lidar com fenómenos colectivos sofisticados (instituições com intencionalidade colectiva), a nossa proposta — podemos reduzir a intencionalidade a mecanismos e esquemas de interacção — continuou aqui a mostrar-se produtiva.

4. Projecto para uma redução heurística da intencionalidade

Nenhum projecto de redução tem sucesso garantido. Por mais poderoso que seja o quadro conceptual que o serve, e por muito pouco que compreendamos a conexão entre a realidade e os nossos conceitos, e mesmo que tenhamos razões filosóficas para desconfiar da própria noção de realidade externa, o certo é que a factualidade vem frequentemente invadir os terrenos do pensamento com perturbações novas obrigando a reformulações conceptuais.

Além do mais, o projecto reducionista clássico suscita hoje dúvidas fundadas — e não apenas as que sempre persistiram entre os estudiosos científicos e filosóficos da biologia. Também da física, o terreno científico que mais abrigo dá tradicionalmente ao reducionismo, vêm poderosos argumentos contrários. Por exemplo, o Prémio Nobel da Física Robert Laughlin, que considera a esperança reducionista uma preguiça intelectual que distrai da diversidade do mundo, demonstrou a existência de estados estáveis da matéria (enquadrados numa categoria designada por "protectorados quânticos", de que são exemplos a supercondutividade e a superfluidez) cuja explicação é independente do nível das partículas elementares, mostrando propriedades genéricas a baixas energias que são insensíveis ao nível microscópico de organização da matéria, sendo determinadas apenas por princípios superiores de organização — o que justifica designá-los como fenómenos físicos colectivos (Laughlin e Pines 2000). Não há razão, pois, para estarmos excessivamente confiantes à partida para qualquer projecto de redução — porque ele não enfrentará o desafio da adequação apenas em termos lógicos, mas sempre também em termos empíricos.

Não obstante, algum sentido terá o facto de nas secções precedentes ter sido possível mostrar a utilidade da hipótese da intencionalidade ser explicável por uma

especificação de mecanismos (organização interna dos itens intervenientes num fenómeno intencional) e um esquema histórico de interacção (estrutura das relações mútuas significativas adquiridas historicamente pelos vários itens intervenientes no mesmo processo intencional)[2]. Em particular, vimos como essa hipótese permite uma compreensão: das semelhanças e diferenças de modos intencionais nos animais e nos humanos; como, no caso das máquinas, a concentração exclusiva no mecanismo prejudicou o projecto da IA; como, no caso da realidade social institucional humana, uma insuficiente atenção à especificação dos mecanismos pode autorizar especulação insustentada.

Podemos dizer, então, que se indicou a possibilidade de um certo tipo de redução da intencionalidade: cada categoria de fenómenos intencionais poderá ser cabalmente explicada por mecanismos e um esquema histórico de interacção. Mas, mesmo assim, tem a intencionalidade de poder ser reduzida de acordo com essa possibilidade?

Sugerimos que sim — pelas razões que justificam designar este projecto como "redução heurística" da intencionalidade. Ela é heurística porque sugere uma estratégia para uma convergência de percursos de investigação. Uma explicação geral da intencionalidade deveria abranger humanos, animais, máquinas e colectivos como sujeitos intencionais; há, a par do trabalho conceptual da filosofia, várias disciplinas científicas que investigam domínios de fenómenos interessantes para a compreensão da intencionalidade em todas aquelas classes de possíveis sujeitos intencionais; algumas dessas linhas de investigação dedicam-se principalmente a estudar os mecanismos envolvidos, enquanto outras seriam mais pertinentes para compreender os esquemas de interacção; existem fronteiras históricas que dificultam a comunicação entre disciplinas que se entendem como tendo objectos muito distintos (sociologia, biologia e física, por exemplo), mas cuja confluência poderia potenciar o avanço na explicação da intencionalidade (por exemplo, aproximando a investigação sobre fenómenos físicos colectivos, fenómenos biológicos colectivos, fenómenos sociais colectivos); a concentração na equação "a intencionalidade é mecanismos mais um esquema de interacção histórica entre esses mecanismos" teria valor heurístico ao criar um foco e uma orientação comum de investigação científica e filosófica — sem eliminar a luxuriante floresta de variedades de intencionalidade.

Um exemplo dos efeitos possíveis dessa redução heurística da intencionalidade seria a mobilização para esta pesquisa de estudos sobre outros fenómenos naturais (além da evolução) que podem explicar o "aspecto interacção". É o caso dos estudos sobre o desenvolvimento pré-natal e pós-natal nos humanos e noutros animais, que investigam como é que, nas espécies que se reproduzem sexualmente, o zigoto, a célula única resultante da fecundação, vem a transformar-se num indivíduo adulto completamente formado. Como é sabido, no caso concreto do desenvolvimento pós-natal nos humanos, esses processos, onde confluem biológico e social,

são responsáveis pela instalação de estruturas básicas quer para controlo do corpo próprio quer para a relação social sofisticada, tornando específico e focado o que à nascença é genérico. Nesse sentido, a compreensão do desenvolvimento contribuirá para a compreensão da intencionalidade enquanto esquema de interacção assente em estruturas naturais.

Outra vertente das vantagens desta proposta de redução heurística da intencionalidade estaria na mobilização de estudos científicos sobre o "aspecto mecanismos", capazes de explicar a aparente "acção a distância" envolvida na intencionalidade. Por exemplo, uma melhor consideração da Hipótese Dinâmica em Ciências Cognitivas (van Gelder 1998), embora carente de depuração de ingenuidades várias, seria talvez capaz de reequilibrar o peso das explicações que dependem de alguma forma da hipótese da linguagem do pensamento (como é o caso até com Millikan), mostrando como num sistema dinâmico (um organismo vivo, por exemplo) um conteúdo representacional individual pode ser inseparável de uma trajectória comportamental (uma micro-história) do próprio sistema. Outro caso que ilustra muito concretamente o que podemos ganhar investigando mais os mecanismos, enquanto componente da intencionalidade, é o dos neurónios-espelho.

Na década de 1990, investigadores que estudavam o córtex motor de macacos, para determinar como os comandos para realizar determinadas acções são codificados por padrões de disparo neuronal, registavam a actividade neuronal quando os macacos realizavam acções como pegar num brinquedo ou alimento. Descobriram então que, quando os macacos viam um humano a pegar nos mesmos objectos, alguns dos seus neurónios disparavam como se eles próprios estivessem a fazer esses gestos. A esses neurónios chamaram neurónios-espelho. Investigações posteriores mostraram que o padrão de actividade neuronal associado à acção observada era uma representação cerebral desse acto, independente do seu autor. E mostraram também que os neurónios-espelho não reagem apenas à acção observada, mas também ao seu significado (regista-se o mesmo tipo de actividade quando os macacos apenas podem observar certas pistas da acção representada, sem acesso perceptivo directo à mesma). Outras experiências indicam que os neurónios-espelho também representarão intenções: codificam diferentemente sequências motoras que, embora expressas por movimentos corporais idênticos (pegar na comida) são realizadas com intenções diferentes (levá-la à boca, guardá-la na caixa). As distinções operadas pelos neurónios-espelho quando o macaco observa são fundadas nas distinções operadas quando ele próprio age. Para várias fases desta investigação foram realizadas experiências com humanos, que parecem indicar que também nós teremos um sistema de neurónios-espelho no nosso cérebro, inclusivamente com ligação às emoções (Rizzolatti *et al.* 2006). O significado científico e filosófico destas descobertas está longe de merecer um consenso estabilizado (Origgi e Sperber 2005). Contudo, estudos como estes podem ajudar a compreender a intencionalidade (e a sua "miste-

riosa" "acção a distância") a partir de mecanismos corporais concretos, dispensando o apelo a propriedades inefáveis.

O estudo cientificamente informado de mecanismos envolvidos na intencionalidade teria ainda a vantagem de dar a parte de razão que cabe às perspectivas internalistas sobre a intencionalidade. É que, como sublinha Fodor (1980), embora do ponto de vista ontológico seja mais informativo saber que Édipo é filho de Jocasta, para prever os comportamentos que levarão à tragédia é mais informativo estar a par de que Édipo não sabe que Jocasta é sua mãe. E parece razoável, ao contar com a parte do mundo na intencionalidade, como querem correctamente as abordagens externalistas, incluir a ignorância dos agentes: que é, sem dúvida, um aspecto do mundo. Um aspecto do mundo que encontraria o seu lugar próprio no "aspecto mecanismo" da proposta de redução heurística da intencionalidade que aqui se defende.

Até porque não podemos querer saber mais acerca da relação entre as linguagens e o mundo do que sabemos acerca do mundo. E a isso responde a estratégia heurística embutida nesta proposta de redução da intencionalidade.

Referências

Agre, P. E. 1997. *Computation and Human Experience*, Cambridge, CUP

Bratman, M. E. 1992. Shared Cooperative Activity. *The Philosophical Review* **101**(2): 327–41.

Brentano, F. C. 1874. *Psychologie vom empirischen Standpunkt*. Leipzig: Duncke & Humblot.

Brooks, R. 1999. *Cambrian Intelligence: the Early History of the New AI*. Cambridge, MA: MIT Press.

Fodor, J. A. 1978. Tom Swift and His Procedural Grandmother. In Fodor 1981, pp. 204–24.

———. 1980. Methodological solipsism considered as a research strategy. In Fodor 1981, pp. 225–53.

Fodor, J. A. 1981. *Representations*. Brighton, Sussex: The Harvester Press.

Gilbert, M. 2007. Searle and Collective Intentions. In S. L. Tsohatzidis (ed.) *Intentional Acts and Institutional Facts*. Dordrecht: Springer, pp. 31–48.

Harnad, S. 1989. Minds, Machines and Searle. *Journal of Theoretical and Experimental Artificial Intelligence* **1**: 5–25.

Harnad, S. 1990, The Symbol Grounding Problem. *Physica* D **42**: 335–46.

Harnad, S. 2002. Symbol Grounding and the Origin of Language. In M. Scheutz (ed.) *Computationalism: New Directions*. Cambridge, MA: MIT Press, pp. 143–58.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.

Laughlin, R. B. & Pines, D. 2000. The Theory of Everything. *Proceedings of the National Academy of Sciences* **97**(1): 28–31.

List, C. 2006. The Discursive Dilemma and Public Reason. *Ethics* **116**: 362–402.

Lukács, G. 1920. Class Consciousness. In G. Lukács, *History and Class Consciousness*, Londres: Merlin Press, 1967 (usada a transcrição disponibilizada em http://www.marxists.org, 21-12-08).

Millikan, R. G. 1984. *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.
———. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
Newell, A. 1980. Physical Symbol Systems. *Cognitive Science* **4**: 135–83.
Newell, A. & Simon, H. A. 1976. Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the Association for Computing Machinery* **19**(3): 113–26.
Nolfi, S. & Floreano, D. 2000. *Evolutionary Robotics*. Cambridge, MA: MIT Press.
Origgi, G. & Sperber, D. (moderadores) 2005, *What Do Mirror Neurons Mean? Theoretical Implications of the Discovery of Mirror Neurons.* (http://www.interdisciplines.org/mirror)
Pettit, Ph. 2003. Groups with minds of their own. In F. Schmitt (ed.) *Socializing Metaphysics*. Londres: Rowman & Littlefield, pp. 167–93.
———. 2007. Rationality, Reasoning and Group Agency. *Dialectica* **61**(4): 495–519.
Putnam, H. 1960, Minds and Machines. Republicação in H. Putnam, *Mind, Language and Reality*. Cambridge: CUP, 1975, pp. 362–85.
Rizzolatti, G., Fogassi, L. & Gallese, V. 2006. Mirrors in the Mind. *Scientific American*, Novembro 2006, pp. 54–61.
Searle, J. R. 1990, Collective Intentions and Actions. In P. Cohen, J. Morgan and M. Pollack (eds.) *Intentions in Communication*. Cambridge, MA: The MIT Press, pp. 401–15.
Searle, J. R. 1995. *The Construction of Social Reality*. New York: The Free Press.
———. 2006. Social ontology: Some Basic Principles. In *Anthropological Theory* **6**(1): 12–29.
Steels, L. 2003. Intelligence with representation. *Philosophical Transactions of the Royal Society* (Mathematical, Physical and Engineering Sciences) **361**(1811): 2381–95.
van Gelder, T. 1998. The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences* **21**: 615–28.
Visscher, P. K. 2003. Dance Language. In V.H. Resh e R. T. Cardé (eds.) *Encyclopedia of Insects*, Academic Press, pp. 284–8.

PORFÍRIO SILVA
Institute for Systems and Robotics
Instituto Superior Técnico - Torre Norte
Universidade Técnica de Lisboa
Av. Rovisco Pais, 1
1049-001 Lisboa
PORTUGAL
porfiriosilva@isr.ist.utl.pt

Resumo. Neste ensaio tentamos uma resposta à seguinte questão: tem a intencionalidade de poder ser reduzida a alguma coisa? Propomos que é possível reduzir qualquer variedade de intencionalidade a uma especificação de mecanismos (organização interna dos itens intervenientes num fenómeno intencional) e um esquema histórico de interacção (estrutura das relações mútuas significativas adquiridas historicamente pelos vários itens intervenientes no mesmo processo intencional). Começamos por esclarecemos o sentido desta proposta a partir da abordagem teleosemântica de Ruth Millikan. Depois procuramos avaliar o interesse e a viabilidade da proposta considerando, sucessivamente, o caso do mundo animal e o caso dos humanos; o caso das máquinas; o caso dos colectivos sofisticados especificamente humanos. Terminamos expondo e defendendo o carácter heurístico da redução proposta.

Principia **14**(2): 255–278 (2010).

Palavras-chave: Intencionalidade, mecanismo, interacção, reducionismo, Ruth G. Millikan.

Notas

[1] É por esta via que esta abordagem resolve o problema do erro representacional, que aflige outras propostas.

[2] Nem todos os esquemas de interacção têm de ser necessariamente históricos, nomeadamente quando são criados de novo. Podem ser explicáveis apenas em termos nomológicos — o que, aliás, justifica a convergência progressiva entre a proposta teleosemântica de Millikan e a explicação causal (por Dretske, designadamente). Apesar disso, mantém-se o requisito de que o aspecto interacção seja dado em termos históricos, porque: primeiro, os aspectos não históricos da interacção devem ser explicáveis em termos de mecanismos; segundo, não será possível eliminar os aspectos históricos da interacção na explicação da intencionalidade; terceiro, os aspectos não históricos da interacção estarão normalmente fortemente interligados com aspectos históricos; quarto, parece necessário combater a tendência para subavaliar a importância dos aspectos históricos na explicação da intencionalidade.