

Proposta de reflexão teórica e análise de padrões conceituais com *data mining*

Álvaro Machado Dias

Universidade Federal de São Paulo

Resumo

Mais do que uma teoria ou modelo, a Teoria da Mente se refere a um campo de estudos voltado à habilidade de se prospectar intenções alheias. Visando contribuir para a discussão teórica e a interpretação da literatura no tema, o presente estudo apresenta: 1. Um mapa conceitual do campo, baseado em *data mining/text mining*; 2. Uma abordagem conceitual inovadora e mais eficiente aos estudos de ToM informacional; 3. Uma discussão crítica da extensão e limites dos principais modelos, baseada na análise da literatura com *data/text mining* e nas perspectivas teóricas anteriormente alinhavadas.

Palavras-chave: teoria da mente; *data mining*; ciência cognitiva; epistemologia.

Abstract

Theoretical discussion and conceptual pattern analysis with data mining. More than just a theory or a model, Theory of Mind represents a field of studies concerned with the ability to prospect someone else's intentions. Aiming to contribute to theoretical discussion and the interpretation of the literature on the matter, this study presents: 1. A conceptual map of the field, based on data mining/text mining techniques; 2. A new and advanced conceptual framework focused on informational ToM studies; 3. A critical discussion of the extensions and limits of the most prominent models, based on the outputs of the data/text mining analysis and on the theoretical perspectives that were previously raised.

Keywords: theory of mind; data mining; cognitive science; epistemology

A prosa silenciosa das mentes

Uma questão importante que perpassa os vários campos de estudos em comunicação humana se refere à extensão e aos limites da possibilidade de se compreender intenções alheias sem a mediação da linguagem. Como é de se considerar, esta capacidade se instaura em um domínio no qual a experiência do sentido se dá de maneira imediata, tal como se por alguns instantes uma mesma paisagem mental perpassasse a experiência de mundo de dois sujeitos. Mas, afinal, como descrever a dinâmica que se abre à 'instalação do outro em meu mundo', e como agregar seus diferentes aspectos de maneira sistemática?

Estas questões, de origem imemorial, compõem o escopo central do campo de estudos da Teoria da Mente, ou simplesmente ToM (de *Theory of Mind*), o qual trata de um conjunto de faculdades cognitivas aparentadas (as diversas maneiras de se prospectar a intenção alheia, tal como descrito abaixo), que podem ser filogeneticamente mais antigas do que a própria linguagem e que permanecem cercadas por grandes dúvidas, as quais se alinhavam no seio de modelos teóricos radicalmente discordantes.

Tal como classicamente proposto por Premack (1988), Teoria da Mente é um conjunto de habilidades mentais que pode ser categoricamente dividido nos tipos perceptivo, motivacional e informacional. Um exemplo clássico de ToM perceptivo é dado

pela capacidade de reconhecer a intenção subjacente às diversas expressões faciais; um exemplo clássico de ToM motivacional dá-se pela propensão em seguir um curso de ação iniciado por outro, ao qual se observa; por fim, ToM informacional se refere à capacidade de compreender e produzir comportamentos eficientes à luz de estados mentais e/ou crenças alheias. O presente artigo tem como objeto de estudo ToM informacional e, mais especificamente, a dicotomia teórica essencial que se desvela no cerne deste campo.

O presente artigo visa: 1. Apresentar um panorama geral do campo, quantitativamente fundamentado (sob fundamentos de revisão sistemática), culminando na produção de um mapa conceitual do campo, pautado por agregados conceituais (*clusters*), componentes de unidades temáticas, representadas no mapa de acordo com a sua importância relativa no campo como um todo. 2. Apresentar uma nova e mais eficiente maneira de se compreender a dicotomia essencial que perpassa o campo de estudos em ToM, através do uso de recursos epistemológicos emprestados da filosofia analítica; 3. Apresentar um balanço acerca da importância relativa das tendências teóricas principais em função dos achados apresentados na seção 2; e, por fim, pontuar as vantagens, desvantagens e limites para a unificação de ambas as posições, da associação entre estes achados e as perspectivas definidas na seção 1.

Método

A primeira seção do artigo introduz um mapa conceitual em 3D (quantitativamente fundamentado) da representatividade relativa dos principais conceitos em voga no campo, associados três a três, fruto do uso de ferramentas de *data mining/text mining* para a extração de padrões da literatura. Para tanto, foi utilizado um sistema semiautomático de extração de conceitos-chave das publicações indexadas no Pubmed e geração de gráficos em 3D. Cabe destacar que a distribuição do campo é feita desde três tesouros (primário, secundário, terciário), os quais contam com centenas de termos, respectivamente hierarquizados à luz de sua representatividade na totalidade das publicações (foram utilizados resumos, títulos e palavras-chave, para evitar vieses). O mapa apresenta como variáveis principais, a altura e o volume de cada pico, os quais refletem a concentração e o número absoluto das publicações elencadas. Mormente, é digno de nota que a apresentação gráfica do mapa busca contribuir para esta percepção (as cores se tornam mais fortes nos cumes), tendo sido inteiramente customizada para tanto. Para uma introdução mais detalhada do método utilizado aqui, ver: Dias (2009a, 2009b, 2010a, 2010b); Dias, Maracaja-Coutinho e Queiroz (2009).

A segunda seção se escora na aplicação inédita do conceito de intencionalidade de John Searle (1984) em ToM, para o delineamento de uma abordagem conceitual capaz de pontuar, de maneira clara e sintética, o que há de essencial entre o investimento cognitivo do sujeito que prospecta intenções alheias e aquele que é fruto desta prospecção e que, por sua vez, prospecta um estado de coisas que se desvela no mundo exterior (em oposição àquele que se desvela no plano das representações mentais).

A terceira seção discute as vantagens e desvantagens de cada posição, considerando sua representatividade no campo. Finalmente, tece considerações sobre a possibilidade de um modelo unificado de ToM.

Resultados

Um mapa conceitual dos estudos em ToM

Uma busca no PubMed através das palavras-chave “Theory of Mind and Human” resgata 2270 publicações (em 08-08-2009), as quais podem ser divididas por tipos de estudos, com representatividades diferentes na totalidade da literatura. Para além desta divisão, é de se ter em vista que os estudos também podem ser separados através de tramas conceituais capazes de expressar o intuito dos pesquisadores. Este tipo de divisão faz-se particularmente estratégica conquanto permite uma resposta aproximada à pergunta: ‘que ideias os pesquisadores que andam publicando neste campo destacam em seus artigos’? Ou seja, esta abordagem permite ir além da divisão metodológica, abrindo as portas da distinção conceitual, a qual é bem mais esclarecedora acerca da ‘infraestrutura’ do campo.

Sob este ponto de partida, apresentamos o mapa topológico abaixo (para detalhes de sua confecção, ver: métodos), o qual partiu de uma organização conceitual do campo em *clusters* de três conceitos.

Como é possível notar, o agrupamento conceitual que mais une pesquisadores é o dos estudos em teoria da mente em crianças, baseado em testes (*task*), sobretudo com enfoque no autismo. Conforme hipotetizamos, isto reflete o fato de que o enfoque primordial em ToM seja o da aplicação do conceito como

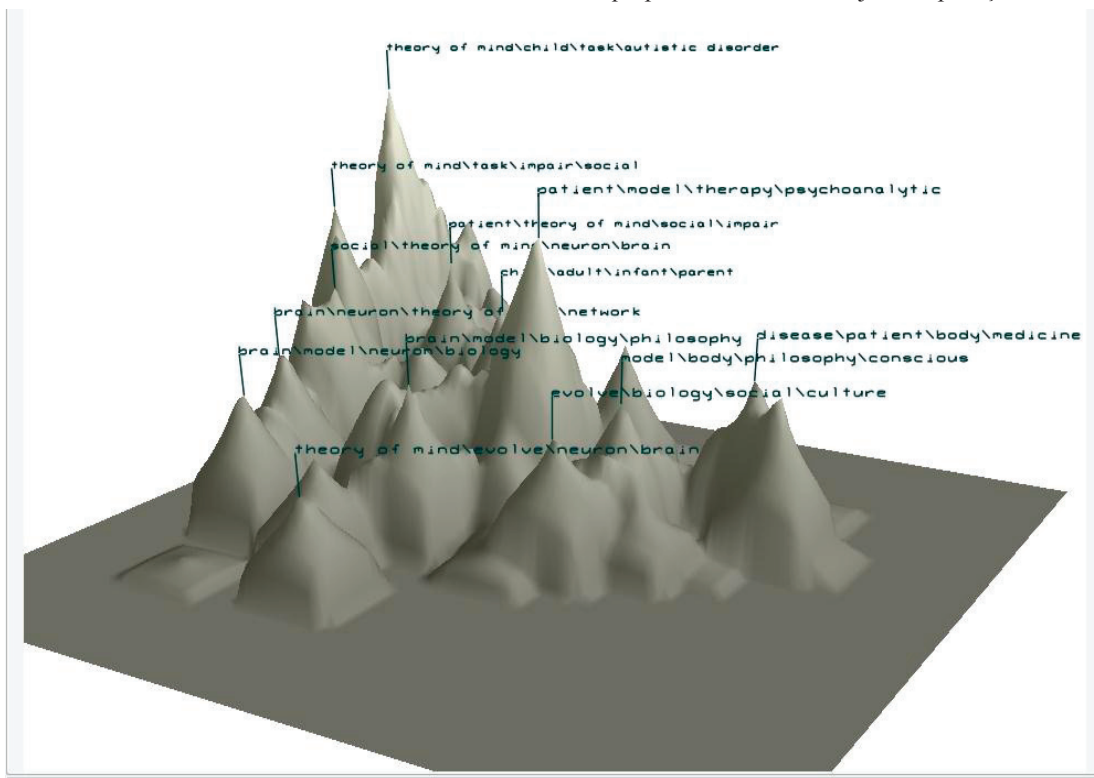


Figura 1. Mapa conceitual dos estudos em ToM indexados no PubMed.

habilidade cognitiva passível de servir de marcador fundamental das alterações cognitivas causadas pelo autismo.

Um segundo agrupamento importante é o que abrange déficits em habilidades sociais de maneira geral (*impaired social*); também se fazem notáveis os agregados em métodos psicoterápicos (*therapy, psychoanalysis*) e, mais dispersamente, diversos envolvendo ‘cérebro’; ‘biologia’; ‘neurônios’ (que neste caso hipotetizamos como referência aos ‘neurônios-espelho’); além de temas caros à filosofia (cultura, consciência, modelos), os quais hipotetizamos como representativos da corrente TT (Theory-Theory) de ToM, tal como veremos a seguir.

Em termos sintéticos, esta organização sugere: 1. Possível divergência teórica de fundo, entre estudos ‘filosóficos’ e clínicos. 2. Predominância destes últimos; 3. Destaque aos estudos e questões relativas à infância; 4. Perspectivas que remetem ao arcabouço biológico da capacidade disseminadas através de enfoques diversos (*networks, brain, neurons, biology*).

Mormente, é possível extrair muitos outros sentidos do mapa, o que esperamos que o leitor faça, adequando tal empenho aos seus interesses particulares.

ToM no Âmbito da Teoria do Conhecimento

Um procedimento analítico para diferenciar sinteticamente a posição do sujeito que prospecta a intenção alheia daquele que é objeto da prospecção. Tendo como ponto de partida esta visão geral do campo que se desvelou através do mapa conceitual, buscaremos definir agora os fundamentos epistemológicos que norteiam as pesquisas no campo. Para tanto, é de se ter em vista que o problema mais elementar que se coloca é o da necessidade de definição de um sentido de intenção (ou intencionalidade) que nos auxilie a descrever eficientemente o processo de apreender o outro na condição de agente intencional, por sua vez subsumida no seio da ideia de lhe prospectar estados mentais. Ainda mais amplamente, esta questão se faz estratégica conquanto uma leitura atenta dos estudos mais famosos do campo revela a existência de divergências explícitas e implícitas associadas aos conceitos de intencionalidade subjacentes.

Doravante, consideremos uma definição desenvolvida especificamente para este contexto e um desenvolvimento (inédito) inspirado no modelo de intencionalidade de Searle (1984): intenção é o aspecto de planejamento e acompanhamento atencional (atenção *covert*, isto é, dada desde o foco mental nos estímulos) de um investimento que altera propriedades de um organismo, do mundo exterior, ou de ambos, e que deixa de existir ao tornar este planejamento e acompanhamento atencional supérfluos. Correlativamente, intenções são particularidades de comportamentos motivados que se esgotam em si mesmas (não se confundem com as alterações do mundo ou de si mesmo produzidas por um ato). Evitando tautologias, acrescentemos: atos motivados são aqueles que ‘aceitam como solução’ (isto é, que aceitam como dispositivo de parada) alterações num estado de coisas (interno/externo), que então se resolve na abolição do planejamento e atenção *covert* que denominamos intenção. Mantidas as definições anteriores, dir-se-ia que a intencionalidade é a propriedade executiva dos organismos que respondem por esta capacidade.

Denominemos por ‘correções’ a estes investimentos

bioenergéticos que geram alterações em estados de coisas internos ou externos; cada correção se caracteriza por um tipo de finalismo: 1. Mente-mundo; 2. Mundo-mente; 3. Mente-mente. Apenas reiterando, destaca-se que, em comum, todos se associam à diminuição do *quantum* informacional (isto é, à atividade cognitiva associada) de que se faz imbuído a execução do comportamento, na medida em que tornam o planejamento e o acompanhamento atencional (tenham eles uma faceta verbal ou não) obsoletos. Da definição de intencionalidade apresentada, segue, *ipso facto*, a existência de tipos distintos de investimento intencional:

1. A energia investida no comportamento de se comer um abacaxi envolve uma correção intencional *mundo-mente*, em que se altera uma condição do mundo (o que faço comendo o abacaxi), adequando-o a um estado mental (volição, comportamento condicionado, etc.), de modo a diminuir o *quantum* informacional determinado no planejamento e acompanhamento executivo para a realização satisfatória da ação;

2. A energia investida para se aprender a jogar gamão, envolve alterar um estado representacional adequando-o ao mundo exterior, isto é, às regras e outras condições de existência do jogo e, portanto, uma correção intencional *mente-mundo*, que igualmente dirime o empenho executivo, que neste caso não está tanto alocado no planejamento, mas, tanto mais no acompanhamento atencional à execução da tarefa;

3. O investimento do romancista em adequar seu conceito do personagem X, para que ele fique mais parecido com o personagem Y, envolve uma correção intencional *mente-mente*, em que se altera um estado interno, para torná-lo mais próximo - ou adequado - a outro; o aspecto informacional é máximo neste caso, tal que a possibilidade de inferi-lo externamente (isto é, sua abertura à prospecção por outrem e, doravante, ToM) torna-se mínima.

Dito isto, podemos passar para uma visão das linhas de força do campo: 1. Os estudos de ToM concentram-se em inferências sobre a possibilidade de prospectarmos correções intencionais *mundo-mente*; 2. Sob este mote, o investimento bioenergético do agente que processa ToM é, por natureza, uma correção *mente-mundo*, na qual o que se entende por mundo alinha-se ao que outrem (aquele que é prospectado) entenderia por *mente*; 3. Os estudos em ToM (de inspiração informacional) se dividem em estudos sobre prospecções de estados mentais emocionalmente imbuídos - isto é, correções intencionais em que o *quantum* informacional a ser reduzido é imbuído de valências e impactação somatossensorial - e prospecções acerca de estados mentais visando objetivos lógicos, isto é, em que a correção intencional do sujeito prospectado dirime um estado informacional rigorosamente executivo, de planejamento e, como tal, não associado à impactação somatossensorial intermediária à sua execução.

Conforme a modelagem de ToM se refere à precisão de mecanismos capacitores da habilidade de se prospectar algo (intenções), decorre o desafio de se determinar o *leitmotiv* e o grau de investimento energético da ação prospectiva, inferencial. Sob este mote, o campo se divide em função de duas concepções, a de que estas ações prospectivas se apresentem como método

de inferência cognitiva e esquemática da realidade mental alheia (denominemos ‘correção *mundo-mente* em sentido forte’, isto é, caracterizada por representações mentais que buscam entender o que se passa na mente de outrem) ou como realidade imanente, dada por equivalência à realidade fenomenológica de outrem (denominemos ‘correção *mundo-mente* em sentido fraco’, dado que esta emergiria sem a necessidade de estímulos buscando emular o que se passa na mente de outrem, no âmbito do ‘teatro da consciência’).

Recapitulando este ponto, desponta o fato de que o campo se define em correntes teóricas dadas pela possibilidade de que a capacidade de perceber as intenções alheias se revele como construção cognitiva motivada pela intenção compreensiva, mediante a qual o agente da inferência viria a alterar sua maneira de pensar, sem se suspender de seus referenciais; ou como percepção, dada na forma de ‘emulação’ da paisagem mental de outrem, naturalmente prescrevendo uma suspensão dos próprios referenciais (ainda que fugaz) e pouco afetando a maneira de se pensar, isto é, o reservatório de constructos cognitivos pertencentes à memória de longo prazo.

Estes modelos são respectivamente denominados TT (*Theory-Theory*) e ST (*Simulation Theory*), sendo digno de nota que, apesar de se autodenominarem teorias, ajustam-se melhor ao conceito de ‘modelos’, como veremos a seguir. Destas considerações igualmente segue que a dicotomia entre estes modelos se reverta em concepções opostas sobre o empenho intencional que se faz necessário para o desencadeamento do ato prospectivo, respectivamente mais planejado e atencionalmente acompanhado em sua execução (modelos TT) ou mais imediatista e sensorial (modelos ST).

Os modelos canônicos de ToM: TT e ST

Modelos TT se baseiam na assunção de que somos dotados de um sistema tácito de axiomas e rotinas (‘leis’) dos quais não temos necessariamente consciência declarativa, mas que nos servem para que infiramos estados intencionais, que assim adquirem o caráter de pequenas teorias que formamos sobre o que se passa na mente de outrem, assim permitindo-nos fazer previsões sobre o seu comportamento. Um dos expoentes desta linha é Paul Churchland (1990, 1991, 1992), que propõe que este conhecimento tácito seja correlato ao que se entende por ‘psicologia do senso comum’, isto é, ao conjunto de crenças e algoritmos para a formação destas, desenvolvidos espontaneamente pela grande maioria das pessoas (e, em contrapartida, não desenvolvido em determinadas circunstâncias psicopatológicas) durante a infância.

Em termos psicogênicos, alguns paralelos com a epistemologia genética piagetiana se fazem pertinentes, à luz da concepção de que esquemas representam aplicativos fundados no conhecimento adquirido sobre o real e no âmbito deste, em grande parte responsável pela possibilidade de experiências compartilhadas (para uma revisão deste ponto, ver: Flavell, 1963; para uma abordagem paradigmática de inspiração TT para ToM na infância, ver: Alison Gopnik, 1992).

Em termos procedurais, fez-se notável a concepção de Donald Davidson (1984) sobre a operação que habilita crenças relativas aos estados intencionais de outrem, que

podemos resumir da seguinte maneira: a capacidade de se operar juízos proposicionais é a chave para a apreensão de intenções e, portanto, para o fenômeno do ToM. Esta depende da possibilidade de conceber estes juízos como verdadeiros ou falsos; segue que alguém é assumido como tendo crenças sobre um estado proposicional de outrem (por exemplo, a afirmação ‘A = X’), conforme constrói uma teoria própria sobre este, em face da qual, a percepção de que o agente da asserção ‘A = X’ atribui ou não veracidade à mesma (a percepção da intencionalidade alheia, ToM), instaura-se dado o imperativo categórico de que este outro sujeito reconhece suas proposições e ações manifestas como verdadeiras ou falsas em função de suas próprias intenções.

Em outras palavras, Davidson acredita que a percepção de estados intencionais alheios emerge da capacidade natural de se perceber a diferença entre “aparência e verdade” em si mesmo. Em um de seus clássicos artigos sobre o tema (Davidson, 1982), ele propõe esta articulação como dispositivo fundante da racionalidade, passível de categorizar bebês e lesmas em oposição a adultos.

Já os modelos de tipo ST se inspiram na máxima de que a capacidade de perceber intenções alheias, bem como outros aspectos da psicologia do senso-comum, pauta-se pela simulação destas em nossas mentes ou, em termos mais amplos, pela capacidade de colocar-se no lugar do outro e com isto ter uma experiência imediata de seus desígnios, não havendo a necessidade de se formar uma teoria (ou projeção de conhecimento) para se entender a intencionalidade alheia; as previsões neste caso seriam endereçadas antes a nós mesmos (ao que Eu faria) do que ao Outro, como no caso anterior. Através dos modelos ST, o escopo de ToM passa a fiar-se à prerrogativa de que a reprodução de um processo em sujeitos distintos gera resultados semelhantes. Esta peculiaridade, associada à maior possibilidade de manipulação experimental, contribuiu para que esta se tornasse a mais difundida categoria de modelos de ToM na atualidade. A despeito do fato de que a origem histórica destes modelos remonta a obras consagradas (Kant, Hume, etc.) é de se considerar que coube a Robert Gordon (1986, 1992a, 1992b) os principais esforços para o seu resgate e reedição contemporânea.

À luz da definição de Gordon e assumindo ‘estímulos cognitivos’ como conceito representativo de todos os tipos de processos cerebrais que atingem a realidade fenomenológica, podemos definir um termo geral de ST: segundo estes, ToM é dado pela suspensão na dinâmica dos processamentos mentais diretamente concernentes ao escopo intencional do Eu, em prol de uma alimentação representativa dos estímulos cognitivos pertinentes ao escopo intencional de outrem e a seus desdobramentos. É justamente esta suspensão de si mesmo, responsável pela possibilidade de ‘emular o outro’ que dirime a necessidade de investimentos intencionais extensivos (investimentos para a correção *mente-mundo* em sentido forte)¹.

A paradigmática capacidade de perceber falsas crenças

Sob a máxima (popperiana) de que as teorias científicas precisam ser falseáveis para que se tornem legítimas, experimentos baseados na formação do conceito de ‘falsas crenças’ se tornaram célebres, conforme propiciaram uma das

mais significativas chaves para se inferir a máxima de que ‘teorias da mente’ são teorias - tal como Gopnik (1992) intitulou um artigo clássico. De maneira geral, estes estudos representam reedições de um clássico experimento de Wimmer e Perner (1983): ‘X’ colocou um chocolate dentro do armário da cozinha; enquanto ele estava ausente, ‘Y’ pegou o chocolate e o utilizou em um bolo, que foi parar na geladeira; daí questiona-se: aonde o garoto irá procurar seu chocolate. Ou então: o sujeito ‘A’ deixa um objeto em ‘X’ e sai; sem que ‘A’ saiba, ‘B’ transfere-o para ‘Y’; a pergunta feita à criança é: aonde ‘A’ irá procurar seu objeto? (Baron-Cohen, et al., 1985).

O que está em jogo com estes experimentos e outros similares não é tanto o delineamento de circunstâncias em que ToM falha, mas a caracterização de circunstâncias propícias para se testar o potencial factível dos modelos TT, já que apenas por haver a formação de teorias é que se faz possível a formação de uma falsa teoria (para uma discussão deste ponto, ver: Davies & Stone, 1998).

Desde este ponto, desenvolveu-se o debate acerca da idade mínima para a ocorrência do fenômeno e então, do estatuto de sua origem, nativista (inerente à mente, inato) ou fruto do aprendizado. Segundo os autores do experimento original de Wimmer e Perner (1983), a idade crítica para a aquisição da capacidade de apontar que o garoto iria procurar o bolo no local onde ele não está mais é por volta dos 4-6 anos (57% de acerto). Fodor (1992) por sua vez, enfatizou que estes experimentos opõem fatos (onde o objeto efetivamente está) a estados proposicionais hipotéticos, o que implicaria considerável capacidade de operar conceitos, a qual teria pouco a ver com a posse de ToM. Assim ele sugeriu que, uma vez eliminados estes entraves metodológicos, evidenciar-se-ia que a aquisição desta capacidade revelar-se-ia ainda mais precocemente e de maneira sugestivamente nativista.

Respondendo a Fodor (1992), Wimmer e Weichbold (1994) afirmaram não encontrar suporte experimental para esta alegação, ao passo que estudos recentes sugeriram que a capacidade de perceber falsas crenças poderia se fiar a um limite tácito em torno dos 15 meses de idade (Onishi & Baillargeon, 2005; Onishi, Baillargeon, & Leslie, 2007); para uma revisão, ver: Perner e Ruffman (2005). Por fim, um estudo clássico de Riggs e Simpson (2005) revelou que a dificuldade de prever falsas crenças se alinha à dificuldade em prever crenças verdadeiras (até cerca dos 4 anos de idade), sugerindo uma associação para com o caráter conceitual da capacidade de compreender crenças, em detrimento da perspectiva nativista relativa à possibilidade desta ser uma faculdade cognitiva irreduzível a outras.

A derradeira conclusão a ser tirada é a de que os estudos em falsas crenças representam uma forte evidência da possibilidade de que, ao menos em certas circunstâncias, ToM seja processado por modelos TT, mas que isto não garante a irreduzibilidade dos modelos TT, nem tão pouco seu caráter nativista.

Arquiteturas de neurônios-espelho como evidências de ToM

O conceito de neurônios-espelho se desenvolveu desde o estudo seminal de Pellegrino, Fadiga, Fogassi, Gallese e Rizzolatti (1992), que reportaram que as células da porção

rostral inferior do córtex pré-motor (F5) de macacos (*Macaca nemestrina*), as quais dispararam sincronizadamente ao longo do processo de inicialização de movimentos voltados a objetivos (isto é, que se associam a ações intencionais no sentido em que definimos anteriormente), também dispararam enquanto os animais observam humanos realizando ações de mesma natureza. Esta descoberta levou os pesquisadores a propor que este padrão de atividade neural estaria coadunado ao sentido da ação e não simplesmente às propriedades do estímulo motor, dando início às pesquisas que buscaram vincular os mais variados modos de interpretação e antecipação de atividades cognitivas e motoras alheias (ToM) a neurônios cujos potenciais de ação fossem sincrônicos aos daquele que sofre a inferência. Subsequentemente este achado foi estendido à porção inferior do córtex parietal de primatas (Rizzolatti & Craighero, 2004), enquanto modelos de arquitetura de neurônios-espelho foram propostos em relação à mente humana, levando à perspectiva de que a arquitetura segue o padrão presente nos outros primatas, sobretudo nas áreas de Brodmann 44 e 46 (Cheng et al., 2009; Cheng, Meltzoff, & Decety, 2007; Gazzola, Aziz-Zadeh, & Keysers, 2006).

Sob esta prerrogativa, os estudos relacionando neurônios-espelho e ToM dividiram-se em: 1. Mecanismos para a compreensão do sentido de ações (Buccino et al., 2001; Buccino, Binkofski, & Riggio, 2004; Catmur, Walsh, & Heyes, 2007; Ferrari, Gallese, Rizzolatti, & Fogassi, 2003; Ferrari, Rozzi, & Fogassi, 2005; Gazzola, Rizzolatti, Wicker, & Keysers, 2007; Keysers et al., 2003; Pellegrino et al., 1992); 2. Imitação, antecipação de ações alheias e empatia (Arbib, Billard, Iacoboni, & Oztop, 2000; Catmur et al., 2007; Iacoboni, Oztop, Kawato, & Arbib, 2006; Rizzolatti, 2005; Rizzolatti, Fadiga, Fogassi, & Gallese, 1999; Williams et al., 2006; Williams, Whiten, Suddendorf, & Perrett, 2001); 3. Inferências diretamente endereçadas a representações intencionais (Dapretto et al., 2005; Iacoboni et al., 2005; Rizzolatti, Fogassi, & Gallese, 2000); 4. Relações entre distúrbios neuropsicológicos, ToM e neurônios-espelho (Dapretto et al., 2005; Fabbri-Destro, Cattaneo, Boria, & Rizzolatti, 2009; Hadjikhani, Joseph, Snyder, & Tager-Flusberg, 2006; Oberman et al., 2005; Oberman & Ramachandran, 2007; Williams et al., 2001).

A descoberta destes neurônios fez-se de sobremaneira reforçadora dos modelos ST posto que, a princípio, tais células dariam escopo a correções *mente-mundo* de tipo imediato e em primeira pessoa, além de, é claro, terem estatuto nativista. Em um estudo clássico, Gallese e Goldman (1998) vão mais adiante sugerindo a possibilidade de que, sob o mote dos neurônios-espelho, o modelo ST não se aplicaria apenas aos humanos, mas ensejaria uma passagem linear entre primatas superiores e nós, no que concerne à apreensão da intencionalidade alheia.

Conclusão: extensão e limites dos modelos TT e ST de ToM

Se tomássemos os estudos publicados ao longo da história para definir qual o melhor modelo de ToM, certamente a opção ficaria com ST. Como não poderia deixar de ser, esta conjuntura surge junto a perspectivas de caráter experimental, extremamente

significativas. Por exemplo, em um estudo visando à avaliação da implicação dos neurônios-espelho para ToM (Iacoboni et al., 2005), avaliou-se, por ressonância magnética funcional (RMf), o grau de ativação da arquitetura de neurônios-espelho em três condições: ação contextualizada; ação descontextualizada; só o contexto sem a ação. Desta forma, revelou-se que ações contextualizadas produzem mais fluxo sanguíneo oxigenado nas áreas comumente relacionadas a neurônios-espelho (por exemplo, BA 44), tal sendo sugestivo de que os mesmos não servem simplesmente à replicação mental de ações observadas, mas ao processamento mental do entendimento de ações voltadas a objetivos e, portanto, de ações intencionais. Outro tipo de perspectiva favorável aos modelos ST é a de que maus funcionamentos em neurônios-espelho produzam uma barreira à sociabilidade, que por sua vez se traduziria em isolamento e solipsismo (para uma meta-análise de estudos recentes, ver: Di Martino et al., 2009).

Dito isto, desponta também o fato de que a compreensão de intenções alheias é algo inexoravelmente diverso da condição de se estar imbuído de intenções originalmente pessoais. Esta é uma conjuntura que não vem sendo explorada na literatura, mas que devemos ter em conta: em se tratando de correções *mente-mundo* em que o que se entende por *mundo* é aquilo que outrem entende por *mente*, estabelece-se sempre alguma distância fenomênica, a qual podemos precisamente caracterizar pela perspectiva de que há um estado informacional a ser reduzido na mente do agente da prospecção que é dado pela satisfação das condições de existência da inferência acerca das intenções alheias, o qual difere do estado a ser satisfeito por correções intencionais de caráter material, ainda que, em ambos os casos, tratem-se de correções *mente-mundo*.

Em outras palavras, ToM não pode ser concebido como uma correção *mente-mundo* convencional, mas antes como uma que produz como *output* a satisfação de um estado de investimento bioenergético associado a uma realidade meramente representacional. Retomando o exemplo apresentado originalmente para a definição de correção *mente-mundo*, não há um abacaxi para pontuar o caráter ontológico do investimento intencional do agente da prospecção; sem maiores desenvolvimentos sugerir-se-ia que o que surge à mente do agente da prospecção é tão somente um estado intencional que, sendo ou não ‘de outrem’, em nada se diferencia, *per se*, dos seus próprios. Do que segue que modelos ST exigem, no mínimo, uma etapa complementar para explicar porque o agente da prospecção não confunde a realidade intencional de outrem com a sua.

ToM exige distanciamento e este é uma medida da construção de um modelo e, portanto, indício de que ao menos uma etapa de determinados tipos de prospecção intencional tenha caráter TT. Sendo um fato digno de nota que a não ocorrência desta distância resulta em delírios de despersonalização, com perfil esquizofreniforme, tal como classicamente descrito por Schneider (1959).

Com isto em vista, mas mantendo a prerrogativa de que os avanços em arquiteturas ST representam saber consolidado, devemos interpretar a máxima enunciada acima como uma sugestão de que a própria função/definição de arquitetura de ST baseada em neurônios-espelho precisa ser revisada no sentido

de abarcar a possibilidade de se discriminar intenções próprias de alheias, através de uma expansão do modelo em direção a arquiteturas mais complexas.

Como hipótese para um desenvolvimento deste tipo, desponta o modelo de *corollary discharge* (Ford, Gray, Faustman, Heinks, & Mathalon, 2005; Ford & Mathalon, 2004, 2005; Ford, Roach, Faustman, & Mathalon, 2008; Hallett, 2007; Mackay, Wulf, Yin, & Abrams, 1993; Poulet & Hedwig, 2007), cuja pedra de toque é a hipótese de que *outputs* motores sejam transmitidos a áreas sensoriais por vias curtas e de maneira opaca à consciência, produzindo um sutil rebaixamento da sensibilidade (da fenomenologia da propriocepção) e abrindo a possibilidade de diferenciarmos naturalmente estímulos produzidos por fontes externas de estímulos proprioceptivos (atenuados). Isto explicaria porque não conseguimos produzir cócegas em nós mesmos, bem como porque não temos a impressão de escutar vozes enquanto pensamos (subvocalizamos), ao contrário da maioria dos portadores de esquizofrenia (Feinberg & Guazzelli, 1999; Ford et al., 2008; Mathalon & Ford, 2008).

Por fim, é preciso se ter em vista a eventual manobra retórica levada a cabo pelos proponentes dos modelos ST ao desviar o foco da intencionalidade para modalidades extremamente simples. Frente a estas, é sintomático retomar o estudo póstumo de Maurice Merleau-Ponty (1969/2002) com o qual introduzimos esta pesquisa, no qual o autor aborda a intencionalidade sem sequer se lembrar de capacidades como a de prever atos motores triviais. É sugestivo que isto decorra da importância desproporcional atribuída às intenções complexas e, sobretudo, àquelas que correm sob o plano da fala (para uma revisão acerca das limitações dos modelos ST em dar conta de ToM ao longo da interpretação da fala, ver: Lotto, Hickok, & Holt, 2009).

Quando nos propomos a compreender o que alguém tem em mente, não buscamos apenas a compreensão dada na forma de um objeto representacional, mas também (e tanto mais) buscamos aferir aonde outrem quer chegar com isto (para uma discussão deste ponto, no domínio da Teoria do Conhecimento, ver: Haye, 2008). Mais profundamente: não nos importa apenas o caráter temático e substancial da correção intencional aferida, mas a sua direção. É precisamente esta direção que não aceita reduções a espelhamentos, posto que estes, por natureza, referem-se apenas a conteúdos manifestos.

Assim é que se insinua a questão: seria o caso da proposição de uma teoria unificadora de ToM? Considerando que, em geral, uma teoria unificadora busca ser reducionista, através da definição de circunstâncias em que um tipo de habilidade é utilizado e circunstâncias em que seria utilizado outro, a resposta deve ser ‘não’. Isto porque tal concepção vai contra uma tendência muito mais atraente (e, infelizmente, menos reducionista e elegante) que é a de se considerar que determinados empenhos cognitivos, passíveis de serem realizados de uma determinada maneira, podem ser realizados de outro modo, eventualmente com algumas perdas de aproveitamento. Neste sentido, seria tanto mais o caso de se considerar uma teoria integrativa, dada pela associação entre ‘modelos de ToM em contiguidade com tipos de empenhos prospectivos’ e ‘sobreposições diversas em vários domínios da vida mental’. Em alguns destes domínios, fazer uso de sistemas ST possivelmente resulta em comportamentos mais

eficientes (por exemplo, conjunturas em que a velocidade de resposta é determinante e a diferenciação entre intenção própria e alheia, secundária), enquanto em outros possivelmente ocorre o inverso (por exemplo, quando a previsão passa por aspectos sutis dos desígnios alheios, tal como a prospecção do ‘não-dito’ mencionado na introdução e as intenções por trás de discursos manifestos de caráter complexo e direções inconscientes).

Considerando ainda que a aplicação de um modelo não-ideal de ToM em determinadas circunstâncias pode resultar em perdas de desempenho temporal ou preditivo, segue que uma nova chave para a compreensão de déficits clínicos e subclínicos de ToM deve emergir do desenvolvimento de um teste capaz de determinar, em domínios diversos, a relação ‘tipo de prospecção utilizada/tipo de prospecção mais eficiente’.

Referências

- Alison Gopnik, H. M. W. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7(1-2), 145-171.
- Arbib, M., Billard, A., Iacoboni, M., & Oztop, E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Networks*, 13(8-9), 975-997.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37-46.
- Buccino, G., Binkofski, F., Fink, G., Fadiga, L., Fogassi, L., Gallese, V., ... Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13(2), 400-404.
- Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language*, 89(2), 370-376.
- Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current Biology*, 17(17), 1527-1531.
- Cheng, Y., Chou, K. H., Decety, J., Chen, I. Y., Hung, D., Tzeng, O. J. L., & Lin, C.-P. (2009). Sex differences in the neuroanatomy of human mirror-neuron system: a voxel-based morphometric investigation. *Neuroscience*, 158(2), 713-720.
- Cheng, Y., Meltzoff, A. N., & Decety, J. (2007). Motivation modulates the activity of the human mirror-neuron system. *Cerebral Cortex*, 17(8), 1979-1986.
- Churchland, P. M. (1990). Eliminative materialism and the propositional attitudes. In W. G. Lycan (Org.), *Mind and cognition* (pp. 206-223). Oxford: Blackwell.
- Churchland, P. M. (1991). Folk psychology and the explanation of human behavior. In J. D. Greenwood (Org.), *The future of folk psychology* (pp. 51-69). Cambridge: Cambridge University Press.
- Churchland, P. M. (1992). *Matter and consciousness*. Cambridge: MIT Press.
- Dapretto, M., Davies, M., Pfeifer, J., Scott, A., Sigman, M., Bookheimer, S., & Iacoboni, M. (2005). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, 9, 28-30.
- Davidson, D. (1982). Rational animals. *Dialectica*, 36(4), 317-327.
- Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Oxford University Press.
- Davies, M., & Stone, T. (1998). Folk psychology and mental simulation. In A. O'Hear (Org.), *Contemporary issues in the philosophy of mind* (pp. 53-82). Cambridge: Cambridge University Press.
- Di Martino, A., Ross, K., Uddin, L. Q., Sklar, A. B., Castellanos, F. X., & Milham, M. P. (2009). Functional brain correlates of social and nonsocial processes in autism spectrum disorders: an activation likelihood estimation meta-analysis. *Biological Psychiatry*, 65(1), 63-74.
- Dias, A. M. (2009a). A depressão em humanos e outras espécies. *Estudos e Pesquisas em Psicologia*, 14, 178-192.
- Dias, A. M. (2009b). Depression across the species. *Bentham Psychiatry Journal*, 3, 33-38.
- Dias, A. M. (2010a). Modelos e análises computacionais em neurociências: revisão sistemática. *Estudos e Pesquisas em Psicologia*, 10, 530-552.
- Dias, A. M. (2010b). Tendências em arquiteturas cognitivas: redes de pequenos mundos e esquizofrenia. *Arquivos Brasileiros de Psicologia*, 62, 35-48.
- Dias, A. M., Maracaja-Coutinho, V., & Queiroz, A. T. L. (2009) The role of neuregulin 1 in schizophrenia: a bioinformatics approach. *Nature Precedings*, 1, 1-8.
- Fabbri-Destro, M., Cattaneo, L., Boria, S., & Rizzolatti, G. (2009). Planning actions in autism. *Experimental Brain Research*, 192(3), 521-525.
- Feinberg, I., & Guazzelli, M. (1999). Schizophrenia - a disorder of the corollary discharge systems that integrate the motor systems of thought with the sensory systems of consciousness. *British Journal of Psychiatry*, 174(Mar), 196-204.
- Ferrari, P., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, 17(8), 1703-1714.
- Ferrari, P., Rozzi, S., & Fogassi, L. (2005). Mirror neurons responding to observation of actions made with tools in monkey ventral premotor cortex. *Journal of Cognitive Neuroscience*, 17(2), 212-226.
- Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. New York: Van Nostrand Reinhold.
- Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition*, 44(3), 283-296.
- Ford, J. M., Gray, M., Faustman, W. O., Heinks, T. H., & Mathalon, D. H. (2005). Reduced gamma-band coherence to distorted feedback during speech when what you say is not what you hear. *International Journal of Psychophysiology*, 57(2), 143-150.
- Ford, J. M., & Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research*, 38(1), 37-46.
- Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *International Journal of Psychophysiology*, 58(2-3), 179-189.
- Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2008). Out-of-synch and out-of-sorts: dysfunction of motor-sensory communication in schizophrenia. *Biological Psychiatry*, 63(8), 736-743.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12), 493-501.
- Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, 16(18), 1824-1829.
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *NeuroImage*, 35(4), 1674-1684.
- Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, 1(2), 158-171.
- Gordon, R. M. (1992a). Reply to Stich and Nichols. *Mind & Language*, 7(1-2), 87-97.
- Gordon, R. M. (1992b). The simulation theory: objections and misconceptions. *Mind & Language*, 7(1-2), 11-34.
- Hadjikhani, N., Joseph, R., Snyder, J., & Tager-Flusberg, H. (2006). Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral Cortex*, 16(9), 1276-1282.
- Hallett, M. (2007). Volitional control of movement: The physiology of free will. *Clinical Neurophysiology*, 118(6), 1179-1192.
- Haye, A. (2008). Living being and speaking being: toward a dialogical approach to intentionality. *Integrative Psychological and Behavioral Science*, 42(2), 157-163.

- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol*, 3(3), e79. doi:10.1371/journal.pbio.0030079.
- Iacoboni, M., Oztop, E., Kawato, M., & Arbib, M. (2006). Mirror neurons and imitation: a computationally guided review. *Neural Networks*, 19(3), 254-271.
- Keysers, C., Kohler, E., Umiltà, M., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, 153(4), 628-636.
- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13(3), 110-114.
- Mackay, D. G., Wulf, G., Yin, C., & Abrams, L. (1993). Relations between word perception and production: new theory and data on the verbal transformation effect. *Journal of Memory and Language*, 32(5), 624-646.
- Mathalon, D. H., & Ford, J. M. (2008). Corollary discharge dysfunction in schizophrenia: evidence for an elemental deficit. *Clin EEG Neurosci*, 39(2), 82-86.
- Merleau-Ponty, M. (1969/2002). *A Prosa do mundo*. São Paulo: Cosac & Naify.
- Oberman, L., Hubbard, E., McCleery, J., Altschuler, E., Ramachandran, V., & Pineda, J. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, 24(2), 190-198.
- Oberman, L., & Ramachandran, V. (2007). The simulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin*, 133(2), 310-327.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255-258.
- Onishi, K. H., Baillargeon, R., & Leslie, A. M. (2007). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica*, 124(1), 106-128.
- Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1), 176-180.
- Perner, J., & Ruffman, T. (2005). Psychology: infants' insight into the mind: how deep? *Science*, 308(5719), 214-216.
- Poulet, J. F. A., & Hedwig, B. (2007). New insights into corollary discharges mediated by identified neural pathways. *Trends in Neurosciences*, 30(1), 14-21.
- Premak, D. (1988). 'Does the chimpanzee have a theory of mind' revisited. In R. W. Byrne & A. Whiten (Orgs.), *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans* (pp. 160-179). New York: Oxford University Press.
- Riggs, K. J., & Simpson, A. (2005). Young children have difficulty ascribing true beliefs. *Developmental Science*, 8(3), F27-30.
- Rizzolatti, G. (2005). The mirror neuron system and imitation. In S. Hurley & N. Chater (Orgs.), *Perspectives on imitation: from Neuroscience to Social Science* (Vol. 1: Mechanisms of imitation and imitation in animals - Social Neuroscience) (pp.55-76). Cambridge, MA: MIT Press.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27(1), 169-192.
- Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1999). Resonance behaviors and mirror neurons. *Archives Italiennes de Biologie*, 137(2), 85-100.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2000). Mirror neurons: intentionality detectors. *International Journal of Psychology*, 35(3-4), 205-205.
- Schneider, K. (1959). *Clinical psychopathology*. New York: Grune & Stratton.
- Searle, J. (1984). *Minds, brains and science: the 1984 Reith lectures*. Cambridge: Harvard University Press.
- Williams, J., Waiter, G., Gilchrist, A., Perrett, D., Murray, A., & Whiten, A. (2006). Neural mechanisms of imitation and 'mirror neuron' functioning in autistic spectrum disorder. *Neuropsychologia*, 44(4), 610-621.
- Williams, J., Whiten, A., Suddendorf, T., & Perrett, D. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehavioral Reviews*, 25(4), 287-295.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128.
- Wimmer, H., & Weichbold, V. (1994). Children's theory of mind: Fodor's heuristics examined. *Cognition*, 53(1), 45-57.

Financiamento FAPESP, Auxílio Regular à Pesquisa 09/54836-7

Notas

1. É de se ter em mente que o modelo de Gordon não é o único modelo ST e que o mantemos em foco por razões didáticas.

Álvaro Machado Dias, Doutor em Neurociências e Comportamento pela Universidade São Paulo e pesquisador do programa de pós-doutorado do Laboratório de Neuroimagem em Psiquiatria "LIM-21" da USP, é professor Adjunto da Universidade Federal do São Paulo. Endereço para correspondência: Rua Dr. Ovidio Pires Campos, s/n, Cerqueira Cesar, São Paulo/SP. CEP: 05403-010. Tel.: (11)3069-8132. E-mail:alvaromd@usp.br e alvaromd@unifesp.br