

Como a teoria da mente e a cognição moral interagem?

How do theory of mind and moral cognition interact?

Ricardo Augusto Perera

Mestrando do Curso de Filosofia da UNISINOS (bolsista CAPES/PROSUP)
Grupo de Pesquisa Social Brains
ricardoperera@outlook.com

Sofia Inês Albornoz Stein

Programa de Pós-Graduação em Filosofia da UNISINOS – Brasil/CNPq
Grupo de Pesquisa Social Brains
siastein@me.com

Resumo: Por “Teoria da Mente” (ToM) entende-se a capacidade de atribuir estados mentais representacionais a outras pessoas, tais como crenças e desejos, e de explicar/predizer comportamentos levando em consideração o modo como esses estados mentais comumente interagem. “Cognição Moral” (CM), num sentido abrangente, refere-se ao modo como pessoas fazem juízos morais, avaliando comportamentos à luz de um sistema de normas. ToM e CM, competências que integram a Cognição Social humana, apesar de se distinguirem em termos computacionais (i.e., quanto a seus papéis funcionais – domínio de inputs que processam e tipos de outputs que produzem) e de realização neural, são processos que interagem de várias formas, havendo tarefas cognitivas que as requerem conjuntamente. Pretende-se, neste trabalho, analisar tipos de tarefas cognitivas nas quais ToM e CM interagem (ToM tomando como input um output da CM e CM tomando como input um output da ToM), assim como nas quais ToM e CM são recrutadas isoladamente. Devido à existência de grupos clínicos (e.g. indivíduos pertencentes ao espectro autista) que apresentam desempenho anômalo em determinadas tarefas cognitivas que recrutam ToM e CM conjuntamente (e.g. julgar o comportamento de um agente que, querendo realizar algo bom, causou um dano acidental a outrem), faz-se necessário investigar o papel exato de cada processo na produção desses outputs mais complexos a fim de ser esclarecida a origem dos erros. Sem levar em conta os resultados conjuntos de experimentos que recrutam uma ToM e uma CM isoladamente, não se pode determinar se o mau desempenho em uma tarefa que dependa do uso coletivo dos dois processos é produto de uma deficiência em ToM, CM ou em ambas.

Palavras-chave: Teoria da Mente. Cognição Moral. Meta-representações. Autismo. Síndrome de Down.

Abstract: “*Theory of Mind*” (ToM) means the ability to attribute representational mental states to others, such as beliefs and desires, and to explain/predict behaviors taking into account how these mental states

commonly interact. “Moral Cognition” (CM), in a broad sense, refers to the way people make moral judgments, evaluating behaviors in the light of a system of norms. ToM and CM, skills that are part of human social cognition, although being distinct in computational terms (i.e. about their functional roles—domain of input processed and types of outputs produced) and neural embodiment, are processes that interact in various forms, with cognitive tasks that require them jointly. It is intended in this paper to analyze types of cognitive tasks in which ToM and CM interact (ToM taking as input an output of CM and CM taking as input an output of ToM), and in which ToM and CM are recruited separately. Due to the existence of clinical groups (e.g. individuals in the autism spectrum) having anomalous performance in certain cognitive tasks that recruit ToM and CM together (e.g. judging the behavior of an agent who, wanting to accomplish something good, caused an accidental injury to another person), it is necessary to investigate the exact role of each process in the production of these complex outputs to clarify the origin of the errors. Without taking into account experimental results that recruit ToM and CM alone it is not possible to determine whether the poor performance on a task that depends on the collective use of the two processes is the result of an impairment in ToM, CM or both.

Keywords: *Theory of Mind. Moral Cognition. Meta-representations. Autism. Down syndrome.*

Introdução

Não obstante o fato de serem ambas integrantes da Cognição Social dos seres humanos, a Teoria da Mente e a Cognição Moral são tipos de processos que diferem em muitos aspectos. “Teoria da Mente” (ToM) refere-se à capacidade de atribuir estados mentais representacionais a outras pessoas, tais como crenças, intenções e desejos, e de explicar/predizer comportamentos levando em consideração o modo como esses estados mentais tipicamente interagem (é comum estas regularidades serem expressas por meio de leis “*Ceteris Paribus*” não-estritas, como “geralmente, se alguém acredita que a ação p produz a consequência q e deseja que q se torne o caso, então, na ausência de meios alternativos a p e desejos incompatíveis com q , esta pessoa faz a ação p ”). A ToM é um processo cognitivo que toma como input “os inputs de outras pessoas” e os estados mentais a elas previamente atribuídos, gerando como output a atribuição de novos estados mentais e/ou comportamentos preditos. A Cognição Moral (CM) — aqui entendida como “o modo como as pessoas processam informações moralmente relevantes para formar juízos morais”¹ —, por sua vez, receberia como input crítico o comportamento de um agente, produzindo, à luz de um sistema de normas, um juízo moral como output. Embora a formação de um juízo moral, tomada integralmente, dependa comumente de várias etapas de processamento de distintas naturezas 2computacionais (e.g. compreensão da sequência de eventos, inferência dos estados mentais do agente, julgamento da ação com base em todas as informações moralmente relevantes que se possui sobre a ação e o agente), por “Cognição Moral” pretende-se fazer referência apenas ao que

1 Adoto a definição de KNOBE, 2005, p. 357.

ocorre na etapa final, na qual são integradas e moralmente analisadas as informações obtidas dos processos precedentes. A Tabela 1 resume, de modo simplificado, o que viria a ser o papel funcional de cada processo:

	Teoria da Mente	Cognição Moral
Domínio de Input	- Inputs de outros agentes - Estados mentais previamente possuídos por outros agentes	- Ações de outros agentes - Estados mentais de outros agentes (output da ToM)
Tipo de Output	- Atribuição de novos estados mentais - Predição de comportamentos	- Juízos morais
Tipo de Processo	- Baseado em regularidades (<i>folk</i>) psicológicas que conectam estímulos, estados mentais e comportamentos	- Baseado em um sistema de normas

Tabela 1 – Os papéis funcionais da Teoria da Mente e da Cognição Moral.

Propõe-se, neste trabalho, analisar como a ToM e a CM, independentes em termos funcionais (e provavelmente sendo também implementadas em regiões cerebrais distintas), são recrutadas em nossas tarefas sociais, buscando exemplos de casos em que apenas a ToM ou a CM são empregadas, assim como em tarefas de maior complexidade em que ToM e CM interagem, permitindo conclusões mais sofisticadas. Procurar-se-á precisamente instâncias dos seguintes casos:

- 1) Tarefas que fazem uso apenas de uma ToM.
- 2) Tarefas em que a CM toma um output da ToM como input.
- 3) Tarefas em que a ToM toma um output da CM como input.
- 4) Tarefas que fazem uso apenas de uma CM.

Devido à falta de clareza sobre a contribuição exata de cada processo cognitivo em certas tarefas experimentais, alguns equívocos podem ser suscitados ao serem analisados seus dados empíricos. Por exemplo, a pressuposição de que juízos morais requerem uma Teoria da Mente pode levar um pesquisador a concluir que, pelo fato de a ToM ser uma capacidade encontrada apenas em humanos, animais não-humanos sejam incapazes de fazer juízos morais, ainda que rudimentares. Porém, se houver uma subclasse de juízos morais que independa da consideração da mentalidade alheia, um tal salto inferencial se revelará precipitado e inválido. Da mesma forma, da baixa performance de indivíduos pertencentes ao espectro autista

em uma tarefa que requer a formação de juízos morais não se poderá concluir imediatamente que haja algo de errado com a Cognição Moral de autistas. Se a tarefa exigia, conjuntamente com a valoração moral, a consideração da mentalidade de um agente, o mau desempenho pode ser integralmente explicado pelo já documentado déficit de autistas em Teoria da Mente, como no caso dos *False Belief Tasks*. E essa mesma explicação pode vir a ser corroborada caso os mesmos indivíduos demonstrem resultados semelhantes aos de neurotípicos em tarefas que façam uso apenas de uma CM.

1 Tarefas que fazem uso apenas de uma ToM

O principal exemplo de tarefa que faria uso de uma Teoria da Mente dissociada da Cognição Moral é o *False Belief Task*. O *False Belief Task* (FBT), criado por Wimmer e Perner (1983), é um teste originalmente concebido para examinar a habilidade de uma pessoa em atribuir estados mentais que envolvam uma representação da realidade. O teste consiste em contar uma história cujo estímulo permite a atribuição de uma crença em conflito com a realidade. Fora inicialmente assumido por muitos pesquisadores que seria necessário, para passar no teste, levar em consideração a crença falsa de um dos personagens, requerendo igualmente uma apreciação de seu papel causal. De acordo com essa visão, seria correto inferir que, se um ser humano ou animal passar no teste, ele possui uma Teoria da Mente.

A versão de FBT exposta na figura 1 apresenta um cenário com duas personagens, Sally (de vermelho) e Anne (de azul), duas caixas (A e B) e um objeto (uma bola) de posse de Sally (1-2). Apenas o sujeito que realiza o teste consegue ver o interior de cada caixa, sendo este vedado às duas personagens. Sally coloca sua bola na caixa A (3-4) e em seguida sai da sala (5-6). Anne, na ausência de Sally, caminha em direção à caixa A, pega a bola e a transfere para a caixa B, caminhando em seguida ao local em que estivera previamente (7-11). Sally retorna e encontra Anne no mesmo local em que estava quando Sally havia saído (9-11). O teste informa que Sally retornou à sala para pegar sua bola, e em seguida pergunta em direção a qual local (A ou B) será o próximo movimento de Sally. A resposta correta será “em direção à caixa A”, pois Sally não viu a bola ser trocada de lugar, possuindo a crença falsa de que a bola permanece em seu local original.

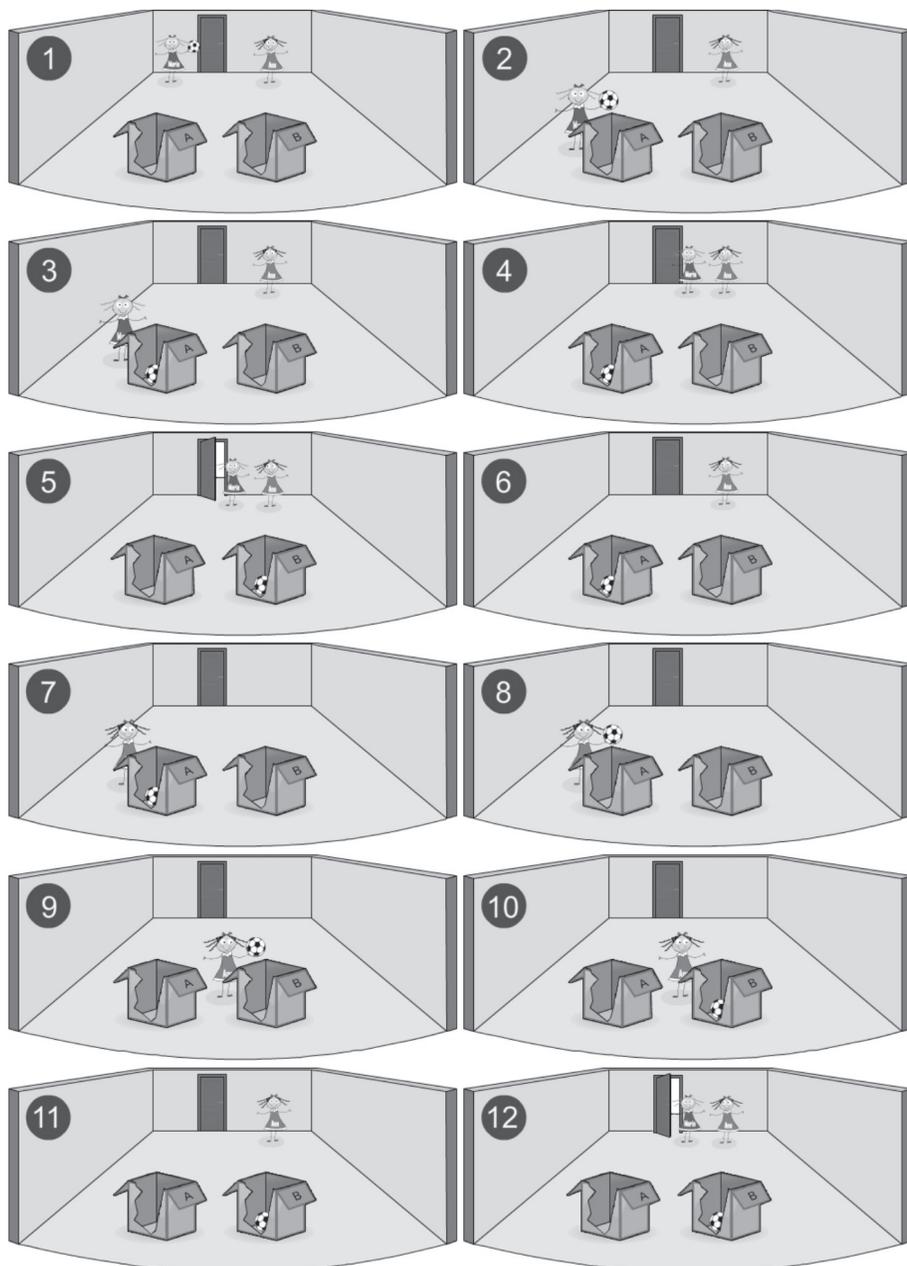


Figura 1 - Sequência do estímulo do *False Belief Task*.

Simon Baron-Cohen, Alan Leslie e Uta Frith (1985), utilizando o mesmo paradigma, compararam a performance de crianças neurotípicas em idade pré-escolar com crianças mais velhas com síndrome de Down e com autismo. Um subgrupo altamente funcional de crianças autistas foi selecionado para o teste, sendo que a média de QI do grupo era de 82 (variando de 70 a 108). A média de QI das crianças com síndrome de Down, por sua vez, era de 64 (variando de 42 a 89). O resultado das crianças neurotípicas e das crianças com síndrome de Down foram surpreendentemente semelhantes: 85% das crianças neurotípicas (23 de 27) e 86% das crianças com síndrome de Down (12 de 14) passaram no teste. Em contraste a esses desempenhos, apenas 20% das crianças autistas acertaram a resposta. Como só havia duas alternativas de resposta, a escolha aleatória geraria 50% de acertos em um grupo que tal prática adotasse, o que indica que crianças autistas optaram preferencialmente pela resposta incorreta. O experimento forneceu fortes evidências de que a habilidade de refletir sobre estados mentais (ou pelo menos um aspecto crucial dela) está dissociada da inteligência geral de um indivíduo, tal como esta é medida em testes de QI: crianças com síndrome de Down, com QI bastante inferior ao das crianças neurotípicas, tiveram um desempenho igual ao destas. Já crianças autistas, cujo QI era superior ao grupo com síndrome de Down, apresentaram uma porcentagem expressivamente maior de erros. Os autores concluíram que o autismo estaria associado a um déficit em uma capacidade de domínio-específico dedicada à leitura de mentes.

Clements e Perner (1994) desenharam uma versão não-verbal de FBT que testaria a Teoria da Mente implícita das crianças. Não obstante o fato de crianças com 3 anos tipicamente errarem ao responder verbalmente onde Sally procurará seu objeto, elas antecipam sua ação olhando para o local em direção ao qual elas acreditam que Sally irá se mover, sem precisarem ser questionadas. Rebecca Saxe (2012) detalha quais são as premissas assumidas nesses testes não-verbais:

Antes que crianças possam falar ou seguir instruções, seus movimentos oculares já revelam suas expectativas acerca do mundo. Crianças olham por mais tempo para objetos ou eventos que as surpreendem e irão apresentar olhar antecipatório em direção ao local em que esperam que algo interessante ocorra (SAXE, 2013, p. 108).

Recentemente, o uso de novos métodos e equipamentos permitiu reduzir ainda mais a idade das crianças que apresentam sucesso preditivo no FBT implícito. Southgate et al. (2007), utilizando um aparelho de *eyetracking*, conseguiram verificar que crianças de 2 anos já apresentariam olhar antecipatório compatível com a atribuição de uma crença falsa a Sally. Onishi e Baillargeon (2005), utilizando o método da “violação da expectativa”, também encontraram evidências que, segundo eles, seriam favoráveis à hipótese de que crianças de 15 meses já possuiriam uma Teoria da Mente representacional (ONISHI; BAILLARGEON, 2005, p. 257). Assumindo-se que crianças olham por mais tempo para eventos que contrariam suas expectativas, as mesmas, caso já possuam uma ToM, apresentarão olhar prolongado apenas quando Sally procurar seu objeto no exato local em que ele se encontrar, olhando relativamente menos quando Sally

procurar onde ela erroneamente acredita que ele se encontra (pois se as crianças antecipam corretamente a ação de Sally, elas se espantarão apenas quando o comportamento de Sally contrariar suas expectativas, que será a procura no local onde o objeto de fato estiver). As crianças de 15 meses do experimento de Onishi e Baillargeon comportaram-se conforme o esperado para alguém que já possui uma Teoria da Mente.

Ainda mais recente e surpreendente é o experimento de Southgate e Vernetti (2014) que, adaptando eletroencefalografia e *eye-tracking* a uma versão modificada de *False Belief Task*, obtiveram evidências de que crianças com apenas 6 meses seriam capazes de antecipar comportamentos com base em crenças falsas (pelo menos de um modo inconsciente). Tomando como premissa que a supressão sensório-motora da frequência alfa (*sensorimotor alpha suppression*) indica predição de ações, a antecipação de comportamentos na versão modificada de FBT produziria essa atividade, não se verificando supressão da frequência alfa quando o realizador do FBT não possui nenhuma expectativa relativa ao comportamento da personagem. Southgate e Vernetti verificaram que, no experimento com crianças de 6 meses, as mesmas antecipavam ações/inações em função da crença falsa da personagem (i.e., por meio da consideração de um estado de coisas não-factual tal como representado pela personagem) e não em função do estado de coisas factual acerca do qual as crianças estavam cientes. Esses resultados não indicariam apenas que crianças de 6 meses seriam capazes de levar em consideração os estados mentais representacionais alheios — até mesmo nos casos em que crença e mundo estão em conflito, como em crenças falsas —, mas que elas, em acréscimo a essa capacidade precoce, já teriam a apreciação dos papéis causais das entidades mentais, isto é, estariam cientes das consequências comportamentais típicas de crenças e das relações destas com inputs, outputs e outros estados mentais.

Senju *et al.* (2009), aplicando o mesmo teste de Southgate et al. (2007) em adultos com síndrome de Asperger com QI elevado (média de 115.6), encontraram evidências de que esse grupo, apesar de apresentar bom desempenho no FBT explícito, fracassa no FBT implícito. Frith (2004) considera que o sucesso preditivo de indivíduos com síndrome de Asperger em FBT's explícitos deve-se ao fato de sua inteligência permitir-lhes utilizar inferências lógicas:

Por exemplo, eles [os indivíduos com síndrome de Asperger] podem apresentar sucesso ao entender cenários complexos, como rastrear onde Mary acredita que John acha que ele pode comprar um casaco. Entretanto, apresentar uma teoria da mente explícita não necessariamente implica uma habilidade intuitiva de leitura de mentes (FRITH, 2004, p. 678).

A figura 2 resume os resultados dos experimentos referidos (simplificados na forma dicotômica “sucesso ou fracasso”):

		FBT Implícito	FBT Tradicional
(Soutgate & Vernetti, 2014)	6 meses	✓	✗ (*)
(Onishi & Baillargeon, 2005)	15 meses	✓	✗ (*)
(Wimmer & Perner, 1983)	3 anos	✓ Teimosia (Perner, 1994)	✗
	5 anos	✓ (*)	✓
(Baron-Cohen, Leslie & Frith, 1985)	10 anos - Síndrome de Down (mean QI = 64)	✓ (*)	✓
	11 anos - Autistas (mean QI = 82)	✗ (*)	✗
(Senju, Soutgate, 2012)	Adultos com Síndrome de Asperger (mean QI = 115)	✗	✓

Figura 2 – Resumo dos resultados dos principais experimentos que utilizaram versões do FBT. Os resultados marcados com asterisco não correspondem a nenhum experimento realizado, consistindo apenas em estimativas inferidas dos demais experimentos.

Faz-se de grande importância, ao serem estudadas, em um determinado grupo de indivíduos, capacidades cognitivas que dependam ou interajam com uma ToM, atentar primeiramente para os desempenhos dos mesmos em tarefas que utilizam uma ToM isoladamente (como expostos na figura 2). Deve-se desconfiar dos resultados de experimentos que utilizam indivíduos com uma ToM deficiente e que estudam outras competências que dependam da consideração de outras mentes. Sem esse cuidado, corre-se o risco de “multiplicar deficiências”, inferindo erroneamente o mau funcionamento conjunto de dois ou mais processos cognitivos autônomos, ainda que um deles esteja intacto. As dificuldades que autistas apresentam em Teoria da Mente podem produzir desempenhos anômalos em tarefas em que a CM toma como input um output da ToM (e.g. juízos morais) e em tarefas em que a ToM toma como input um output da CM (e.g. “*Knobe Effect*”) — temas explorados nas seções seguintes.

2 Tarefas em que a CM toma um output da ToM como input

A capacidade de atribuir estados mentais (ToM) tem papel crucial no modo como fazemos certos juízos morais. Não basta, por exemplo, para que julgemos a ação de um indivíduo x, que nos relatem que x colocou veneno no café de y. Informações acerca de sua perspectiva epistêmica são necessárias para irmos

adiante no juízo moral. Se o veneno estava em um recipiente rotulado como açúcar, x pode ter colocado o veneno acreditando estar a colocar meramente açúcar (o que se afiguraria como uma ação moralmente aceitável). O output da ToM, nesse caso — uma crença falsa ou ausência de intenção de envenenar o café — é tomado como input pela Cognição Moral; esta, por sua vez, produz o output de inocentar o agente (ou diminuir sua culpabilidade), conforme exposto na figura 3:

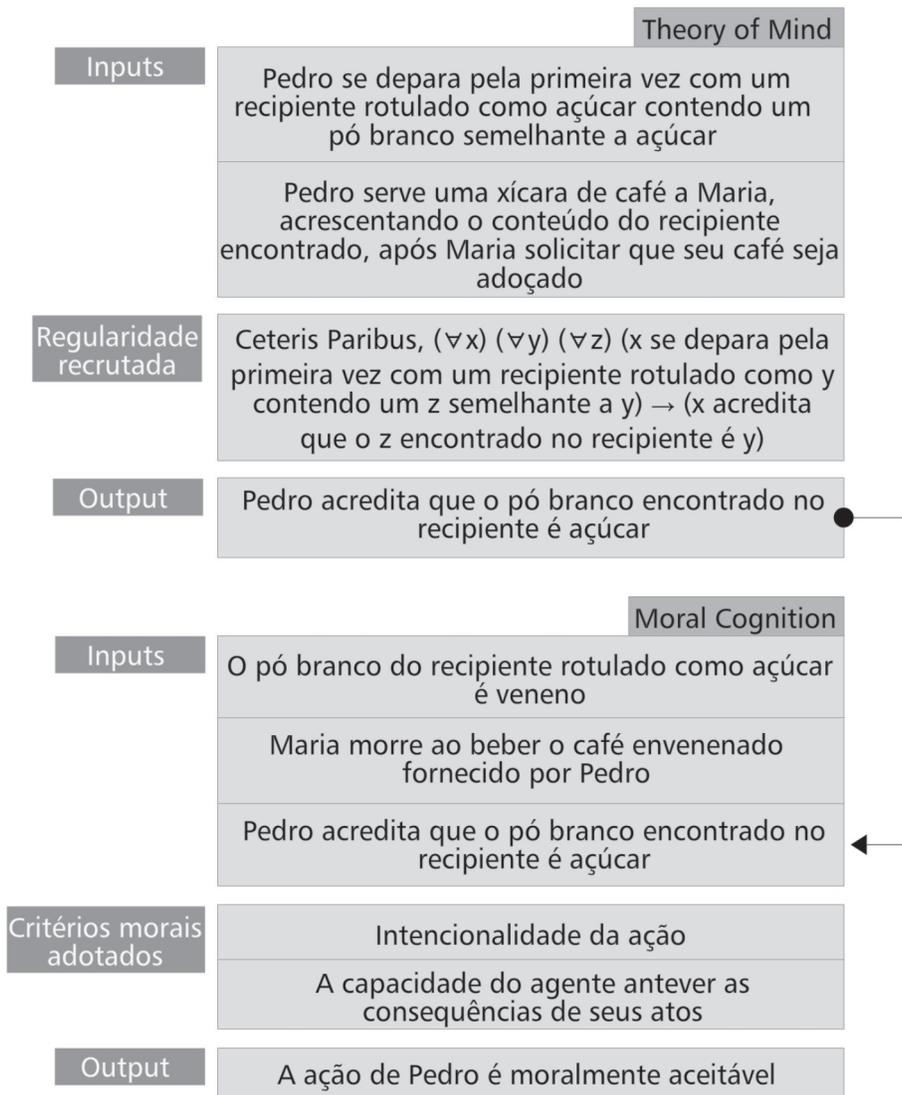


Figura 3 – Exemplo de tarefa em que o que a CM toma um output da ToM como input

Quando indivíduos têm que considerar tanto a ação como a perspectiva epistêmica de um agente para julgar se este se comportou de maneira louvável ou repreensível e interfere-se nas regiões do cérebro associadas à Teoria da Mente por meio de estimulação magnética (*Transcranial Magnetical Stimulation*), os juízos morais são alterados². O peso do fator “estado mental inobservável” acaba sendo atenuado, resultando em juízos enviesados que superestimam o fator “ação observável” (e.g. um agente que, querendo fazer uma boa ação, causa um dano acidental tem o comportamento menos tolerado pelo grupo que está sob influência do TMS, em comparação com o grupo controle, devido à maior dificuldade de consideração de estados mentais). Young e Saxe (2009) também constataram que, das regiões associadas à Teoria da Mente, a Junção Temporoparietal direita, especialmente recrutada quando atribuímos estados mentais representacionais a outras pessoas, apresenta maior ativação em indivíduos que emitem juízos morais mais brandos (i.e., que desculpam mais o comportamento do agente, supostamente por darem mais ênfase à sua mentalidade inocente do que ao seu comportamento acidentalmente nocivo).

O déficit em Teoria da Mente encontrado em indivíduos pertencentes ao espectro autista também refletiria em sua capacidade de julgar comportamentos (em que o fator “estado mental inobservável” também é atenuado na produção de juízos morais).³ O fato de em neurotípicos a interferência nas regiões associadas à ToM prejudicar seus juízos morais é forte evidência de que uma Cognição Moral intacta nem sempre é suficiente para chegarmos a uma valoração adequada de ações. O output de uma ToM afetada compromete todo o processamento ulterior, não obstante a Cognição Moral estar em pleno funcionamento. A ToM permite inferir, a partir da descrição objetiva de uma sequência de eventos envolvendo um agente, os estados mentais que foram causados pela interação com o cenário, assim como os estados mentais que, por sua vez, causaram alterações neste. Dessa forma, é possível extrair novas informações que não estavam presentes na descrição inicial; muitas dessas informações, como as intenções do agente, são cruciais para um juízo moral adequado.

3 Tarefas em que a ToM toma um output da CM como input

Um caso inusitado de interação entre ToM e CM é quando um estado mental é atribuído a um agente em função da aceitabilidade moral de sua ação. Knobe (2003) verificou que as pessoas às vezes consideram uma ação como sendo intencional dependendo de sua significância moral, não ocorrendo este tipo de inferência em ações moralmente neutras que apresentam a mesma estrutura. No experimento de Knobe, participantes liam duas histórias descrevendo o comportamento de um executivo que escolhia iniciar um novo programa que ou prejudicaria (1) ou ajudaria (2) o meio ambiente. O executivo, nas duas condições, afirmaria de modo incisivo “não se importar com as consequências ambientais dos programas”, aprovando ambas as execuções. 85% dos participantes afirmaram que o executivo intencionalmente prejudicou o meio ambiente na primeira condição (dano ambiental), enquanto que

2 YOUNG; CAMPRODON; HAUSER; PASCUAL-LEONE; SAXE, 2010.

3 MORAN; YOUNG; SAXE; LEE; O'YOUNG; MAVROS; GABRIELI, 2011.

na segunda condição (ajuda ambiental) apenas 23% dos participantes disseram que o executivo intencionalmente ajudou o meio ambiente.

O viés cognitivo encontrado por Knobe (“*KnobeEffect*”) entra em conflito com o “*folkconcept*” de ação intencional anteriormente pesquisado pelo próprio autor. Malle e Knobe (1997) encontraram, em um dos primeiros experimentos de Filosofia Experimental, cinco componentes que seriam condições necessárias e suficientes para que um comportamento seja considerado intencional: o desejo do agente pela consequência, crenças sobre a ação em questão que levariam à consequência, a intenção de realizar a ação, consciência (*awareness*) do ato ao realizá-lo e grau suficiente de habilidade para pô-lo em prática.⁴ É racional, partindo desses moldes, negar que o executivo tenha ajudado o meio ambiente intencionalmente, visto que ele próprio dissera ser indiferente às consequências ambientais, isto é, não possuindo o desejo de agir em benefício do meio ambiente. Na ausência de desejo do agente pela consequência, infere-se a ausência de intencionalidade na ação. Mas o executivo também não possuía o desejo de prejudicar o meio ambiente no primeiro caso; entretanto, um juízo de valor (output da Cognição Moral) de certa forma modulou o output da Teoria da Mente, resultando em atribuição de uma intenção. Não está clara a natureza desta interferência. Reações emocionais evocadas por comportamentos moralmente maus poderiam distorcer os julgamentos da Teoria da Mente das pessoas, considerações morais poderiam ter um papel na pragmática de uso de certos termos, mas não na semântica dos conceitos da Teoria da Mente, ou — uma terceira hipótese — as considerações morais de fato têm um papel na competência fundamental subjacente às capacidades da Teoria da Mente.⁵ Adams e Steadman (2004) sustentam que a natureza pragmática da linguagem intencional poderia explicar os achados de Knobe prescindindo da admissão de anomalias semânticas no conceito de ação intencional, uma vez que “dizer que uma ação foi realizada ‘intencionalmente’ ou ‘propositalmente’ implica, pragmaticamente, culpa moral — ainda que a culpa não seja parte do conteúdo semântico literal em ‘S fez A intencionalmente’” (ADAMS; STEADMAN, 2004, p. 269). Se isso for o caso, os participantes não estariam se comprometendo com a existência de um estado mental específico devido às consequências morais de uma ação, mas tão somente culpando o agente por ter danificado o meio ambiente ao expressarem que o mesmo agiu intencionalmente.⁶ Vale enfatizar que o estudo da Cognição Social tem como objetivo compreender “o modo como as pessoas processam estímulos socialmente relevantes”, e não o modo como um agente racional *deveria* processar as mesmas informações a fim de evitar inconsistências internas e interferências de informações irrelevantes. Da mesma forma que em autistas uma ToM deficiente produz juízos morais anômalos, uma Cognição Moral deficiente poderia resultar em atribuições mentalistas diferenciadas (como poderia ser o caso de psicopatas).

4 Cf. GUGLIERMO; MONROE; MALLE, 2009.

5 KNOBE, 2005, p. 358.

6 Knobe (2004), respondendo às objeções de Adams e Steadman (2004), modificou o experimento trocando a palavra “intencionalmente” pelo uso de “razões para agir” (e.g. “O executivo prejudicou/ajudou o ambiente com a finalidade de aumentar os lucros”), obtendo o mesmo efeito do experimento original. O autor argumenta que o novo termo, ao contrário do primeiro, não expressaria uma visão de louvor ou culpa (KNOBE, 2004, p. 184).

4 Tarefas que fazem uso apenas de uma CM

Uma possível maneira de acessar a Cognição Moral de um indivíduo que apresenta déficit em Teoria da Mente seria utilizando tarefas de emissão de juízos morais que excluam a necessidade de atentar para a perspectiva epistêmica de outros agentes. Ao julgarmos um comportamento que instancie a estrutura de um *Trolley Problem* (e.g. um médico que deixa de fornecer um medicamento escasso a um paciente que chegou primeiro, deixando-o morrer, para salvar cinco outros pacientes com um quinto da dose necessária para salvar o primeiro), parece desnecessário levar em consideração os estados mentais do agente. Fatos objetivos de sua ação (de que ele, ao se deparar com o dilema de deixar ou um ou cinco pacientes morrerem, optou pelo menor número de mortes) já permitiriam um julgamento (sob uma perspectiva utilitarista, a ação seria aceitável, pois a alternativa escolhida é a que promove bem-estar ao maior número de pessoas). Se isto for o caso, espera-se — assumindo estar inalterada a Cognição Moral em autistas — que os juízos de autistas de alto desempenho cognitivo não difiram dos de neurotípicos em tarefas como a supramencionada. Para isso, deve-se evitar tarefas que descrevam situações em que as consequências subjetivas, antevistas pelo agente, são diferentes das consequências objetivas da ação, como no exemplo da seção 2 (do café envenenado), em que o agente esperava fazer uma boa ação ao colocar o veneno no café, não podendo seu comportamento ser desculpado apenas pela consideração das consequências, prescindindo de uma ToM.

Considerações Finais

A argumentação aqui exposta foi uma tentativa de avaliar como dois processos cognitivos distintos — Teoria da Mente e Cognição Moral — interagiriam na produção de “conclusões” (= outputs) de maior complexidade. Poder-se-ia dizer que, se a posse de uma ToM for de fato uma característica exclusivamente humana, uma Cognição Social que dela faça uso, combinando-a com outros processos cognitivos (alguns deles possivelmente também unicamente humanos), permitiria inferências ainda mais singulares e em maior grau interditas a animais não-humanos. A procura por tarefas cognitivas que utilizem uma ToM e uma CM isoladamente é uma oportunidade de se ter uma ideia do papel de cada uma delas ao serem posteriormente integradas em processos de níveis superiores. A análise conjunta de experimentos com indivíduos neurotípicos e autistas em Teoria da Mente (por meio de *False Belief Tasks*) e Cognição Moral dependente de uma ToM (e.g. emissão de juízos morais em histórias em que o agente faz um dano acidental) é um exemplo de procedimento dessa natureza. Não só autistas apresentam juízos morais alterados em relação a neurotípicos em circunstâncias normais, como também é possível prejudicar os juízos de neurotípicos ao interferirmos em suas regiões cerebrais associadas à ToM (como a junção temporoparietal direita). Sem levar em conta os resultados conjuntos desses experimentos não se pode determinar se o mau desempenho em uma tarefa que dependa do uso coletivo dos dois processos é produto de uma deficiência em ToM, CM ou em ambas. Da mesma forma que uma subclasse de juízos da CM dependeria de uma ToM, esta também só funcionaria plenamente quando outros processos/competências se mantiverem intactos

(atenção, memória, capacidade de manter em mente representações conflitantes da realidade, linguagem, meta-representações etc.), o que viria a dificultar ainda mais a detecção da origem dos resultados anômalos. Será possível, por exemplo, uma lesão na junção temporoparietal esquerda, dificultando a consideração da perspectiva alheia,⁷ interferir em toda uma cadeia de inferências, como a atribuição de estados mentais, o que, por sua vez, comprometeria avaliações morais? Nesse exemplo, o fracasso em um experimento em que a CM toma como input um output da ToM não se explicaria pela disjunção inclusiva “deficiência na ToM ou na CM”, visto que uma ToM e uma CM intactas ainda não seriam suficientes para julgar ações alheias. O fracasso dependeria de muitos outros processos que poderiam vir a comprometer todas as inferências ulteriores. Se um processo cognitivo complexo X (e.g. a valoração de uma ação humana) depende da plena interação dos sub-processos A, B, C, D e E (e.g. atenção, acesso semântico, meta-representação, Teoria da Mente e Cognição Moral), interferir em regiões cerebrais associadas a A, B, C, D e E pode alterar o output de X. Aplicando o mesmo raciocínio a um caso clínico Y (e.g. autistas e indivíduos com lesões cerebrais), não se poderá determinar, sem analisar o desempenho de Y em tarefas que recrutam A, B, C, D, e E isoladamente, o tipo exato de deficiência que está a causar anomalias no output de X, pois não estará claro qual das etapas inferenciais está comprometendo o raciocínio integralmente.

Referências

- ADAMS, F.; STEADMAN, A. Intentional action and moral considerations: still pragmatic. *Analysis*, 64.3, p. 268-276, 2004.
- APPERLY, I. A. *Mindreaders: The cognitive basis of “Theory of Mind”*. New York: Psychology Press, 2011.
- CALL, J.; TOMASELLO, M. Does the chimpanzee have a Theory of Mind? 30 years later. *Trends in Cognitive Sciences*, May: 12 (5), p. 187-192, 2008.
- CLEMENTS, W. A.; PERNER, J. Implicit understanding of belief. *Cognitive Development*, 9, p. 377-395, 1994.
- FRITH, Uta. Emanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 45: 4, p. 672-686, 2004.
- GOPNIK, A.; WELLMAN, H. M. Why the child’s theory of mind really is a theory. *Mind and Language*, 7 (1-2), p. 145-171, 1992.
- GUGLIERMO, S.; MONROE, A. E.; MALLE, B.F. At the heart of morality lies Folk Psychology. *Inquiry*, 52 (5), p. 449-466, 2009.
- KNOBE, Joshua. Intentional action and side effect in ordinary language. *Analysis*, 63, p. 190-193, 2003.
- KNOBE, Joshua. Intention, intentional action and moral considerations. *Analysis* 64.2, April 2004, p. 181-187, 2004.

7 De acordo com Apperly (2011, p. 77), a junção temporoparietal esquerda estaria associada a um processo de domínio geral de tomada de perspectiva.

KNOBE, Joshua. Theory of mind and moral cognition: exploring the connections. *Trends in Cognitive Sciences*, August 2005, Vol. 9, No. 8, p. 257-359, 2005.

MALLE, B.; KNOBE, J. The folk concept of intentionality. *Journal of Experimental Social Psychology* 33, p. 101-121, 1997.

MORAN, J. M.; YOUNG, L.; SAXE, R.; LEE, S. M.; O'YOUNG, D.; MAVROS, P.; GABRIELI, J. Impaired theory of mind for moral judgment in high-functioning autism. *PNAS*. 108 (7), p. 2688-2692, 2011.

ONISH, K.; BAILLARGEON, R. Do 15-month-old infants understand false beliefs? *Science* 308, p. 255-257, 2005.

POVINELLI, D. J.; VONK, J. Chimpanzee minds: suspiciously human? *Trends in Cognitive Sciences*, Vol. 7, No. 4, p. 157-160, 2003.

PREMACK, D.; WOODRUFF, G. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, p. 515-526, 1978.

SAXE, R.; KANWISHER, N. People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, 19, p. 1835-1842, 2003.

SAXE, R.; POWELL, L. It's the thought that counts: Specific brain regions for one component of Theory of Mind. *Psychological Science*, 17, p. 692-699, 2006.

SAXE, R.; CAREY, S.; KANWISHER, N. Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, 55, p. 87-124, 2004.

SAXE, Rebecca. The new puzzle of Theory of Mind development. In BANAJI, M.; GELMAN, S. (eds.). *Navigating the social world: What infants, children, and other species can teach us*. New York: Oxford University Press, 2013. p. 107-113.

SAXE, Rebecca. The right temporo-parietal junction: a specific brain region for thinking about thoughts. In LESLIE, A.; GERMAN, T. (eds.). *Handbook of Theory of Mind*. Philadelphia: Psychology Press, 2010.

SAXE, Rebecca. Theory of Mind (Neural Basis). In BANKS, W. P. (ed.). *Encyclopedia of Consciousness*. Boston: Elsevier/Academic Press, 2009.

SENJU, Atsushi. Spontaneous theory of mind and its absence in autism spectrum disorders. *The Neuroscientist* 18 (2), p. 108-113, 2012.

SENJU, A. *et al.* Mindblind Eyes: An Absence of Spontaneous Theory of Mind in Asperger Syndrome. *Science* 325, p. 883-885, 2009.

SOUTHGATE, V.; SENJU A. Action anticipation through attribution of false belief by 2 year olds. *Psychological Science* 18 (7), p. 587-592, 2007.

SOUTHGATE, V.; VERNETTI A. Belief-based action prediction in preverbal infants. *Cognition*, 130, p. 1-10, 2014.

YOUNG, L.; SAXE, R. Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychology*, 47(10), p. 2065-2072, 2009.

YOUNG, L., CAMPRODON, J., HAUSER, M., PASCUAL-LEONE, A.; SAXE, R. Disruption of the right temporo-parietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgment. *PNAS*, 107(15), p. 6753-8, 2010.

YOUNG, L.; CUSHMAN, R.; HAUSER, M.; SAXE, R. The neural basis of the interaction between theory of mind and moral judgment. *PNAS*, 104(20), p. 8235-8240, 2007.

Endereço/ Address

Ricardo Augusto Perera
Sofia Inês Albornoz Stein
Laboratório de Filosofia Experimental e Estudos da Cognição – UNISINOS
Av. Unisinos, 950 – Bloco 09 – sala 230
CEP 93.022-750 – Bairro Cristo Rei
São Leopoldo – RS – Brasil

Data de envio: 25-01-16

Data de aprovação: 31-03-16

