

Robôs podem ser moralmente responsabilizados? Reflexões à luz do pragmatismo de John Dewey

Can robots be morally responsible? Reflections from John Dewey's pragmatism

*Fernando Cesar Pilan
Universidade Estadual do Norte do Paraná*

Resumo

Os dois principais objetivos do presente trabalho são: (1) brevemente apresentar os avanços da Inteligência Artificial e (2) problematizar o “dilema moral” que esses avanços geram, buscando contribuições na Teoria Moral de John Dewey. Na esteira dos modelos éticos tradicionais, esse tipo de debate parece não encontrar muitos avanços, pois se assenta na noção de dever moral e imputação de culpa. Nosso objetivo é mostrar que, no viés filosófico pragmatista de Dewey, seria possível uma reflexão sobre a conduta no plano da ação não mais detida no domínio representacionista, no plano do dever. Para Dewey, se um ser é capaz de atualizar seus hábitos, este pode, em algum sentido, ser considerado responsável por suas ações, pois os procedimentos morais não se distinguiriam dos procedimentos cognitivos. Nessa perspectiva, precisaríamos de uma redefinição da noção de responsabilidade, não mais pautada na imputação de culpa, mas na correção de hábitos. Acreditamos que a sugestão de Dewey, de que a conduta seria um processo dinâmico e intimamente ligado à experiência e à cognição, nos oferece ferramentas para pensar alternativamente a questão da responsabilidade no contexto filosófico contemporâneo.

Palavras-chave

Inteligência Artificial; cognição incorporada e situada; responsabilidade; moral; pragmatismo.

Abstract

The two main goals of this work are: (1) briefly to explain how the advances in the Artificial Intelligence has allowed the development of machines with increasingly performances similar to those of living beings, and (2) before these advances, present the moral dilemma generated by them seeking contributions in Moral Theory of John Dewey. In the approach of traditional ethical models this kind of debate does not seem to find many advances, because the moral duty is taken as the starting point of responsible in dualists terms. To Dewey's pragmatism any robots are able to update our habits, they may in some sense be held responsible for their actions: moral procedures are not distinguished from cognitive procedures. We believe that Dewey's suggestion in which the conduct is a dynamic process and deeply linked to the experience offers us tolls to think the question of responsibility alternatively in the contemporary philosophical context.

Keywords

Artificial intelligence; embodied and situated cognition; responsibility; moral; pragmatism.

Introdução

Um dos principais objetivos da chamada “Inteligência Artificial Forte” (IA) nas últimas décadas tem sido desenvolver máquinas capazes de imitar ações e condutas tipicamente de seres vivos. Esse objetivo tem levado a IA a conseguir avanços significativos, principalmente no que se refere ao desenvolvimento de máquinas capazes de agir no mundo de forma cada vez mais sofisticada e autônoma. Um dos problemas que emergem desse panorama atual da IA diz respeito ao âmbito moral destes seres artificialmente desenvolvidos. A partir desse pano de fundo, nossos principais objetivos serão: (1) apresentar uma perspectiva recente em IA – a abordagem da Cognição Incorporada e Situada (C.I.S.) –, considerada um marco

no desenvolvimento de máquinas com condutas e ações cada vez mais próximas às de seres vivos; e (2) mostrar a necessidade de um debate ético não mais pautado nos modelos filosóficos tradicionais, sugerindo possíveis contribuições da teoria da ação contemporânea do pragmatista John Dewey.

A abordagem da Cognição Incorporada e Situada como alternativa à Ciência Cognitiva Tradicional

Por um lado, podemos compreender que a Ciência Cognitiva e a Inteligência Artificial como projetos de pesquisa datam da década de 1950. Nesse sentido, seriam áreas muito recentes que teriam em comum o interesse em compreender a natureza e o funcionamento dos processos cognitivos, com o intuito de construir modelos capazes de imitar nosso comportamento. Por outro lado, podemos também compreender esses dois campos como possuindo uma origem muito mais longínqua. Segundo Georges Vignaux (1991, p. 15),

O desejo de realizar máquinas inteligentes data da antiguidade. E, sem dúvida, uma das primeiras menções do gênero poderia encontrar-se no canto XVIII da Ilíada, no qual Hefesto, deus do fogo, tinha construído mesas com três pés munidas de rodízios, que se deslocavam sozinhas nos palácios dos deuses, uma espécie de robots antes do tempo!

Como exemplos de que a preocupação com a construção de máquinas inteligentes é um anseio humano de longa data, o autor também cita a construção de um autômato chamado *Golem*, feito de madeira e argila, do qual encontramos registro na tradição judaica. Também podemos mencionar a ideia de um “animal-máquina”, proposta pelo mecanicismo cartesiano do século XVII. Enfim, com essas elucidações, Vignaux pretende mostrar que, diferentemente do que muitos afirmam, as preocupações da IA e da Ciência Cognitiva não são produtos do século XX, mas constituem um corpo de problemas relacionados à natureza dos processos cognitivos que inquietam a humanidade há séculos.

Apesar da inquietação com os problemas relacionados à mente poder ter sua origem identificada há séculos, foi somente a partir dos avanços lógico-matemáticos do início do século XX que tivemos o marco de surgimento da IA e da Ciência Cognitiva enquanto áreas de pesquisa. Em primeiro lugar, destacamos que o surgimento da Ciência Cognitiva, conforme definido por Dupuy (1996, p. 9), pode ser datado em 1940, oriundo principalmente

das “Conferências Macy”. Estas foram uma série de conferências levadas a cabo por um grupo de pesquisa formado por lógicos, matemáticos, engenheiros, fisiologistas, psicólogos, antropólogos e economistas que buscavam empreender uma *ciência geral do funcionamento da mente*. Em sua origem enquanto área de pesquisa, a Ciência Cognitiva recebeu o nome de Cibernética, sendo desenvolvida, segundo Dupuy, a partir do choque provocado pelas descobertas lógicas da década de 1930, possibilitando avanço nas áreas de investigação sobre o conhecimento. Destacamos que as referidas descobertas lógicas foram devidas, em parte, às abordagens alternativas surgidas na filosofia contemporânea, que reconheceram a relevância dos processos e fenômenos contingentes no âmbito da investigação filosófica. Um marco que pode ser sublinhado foi a “Virada Naturalista”¹, proveniente do darwinismo e sistematizada pelo pragmatismo de C. S. Peirce, W. James e John Dewey, quando propõe a mudança de um padrão lógico meramente formal, fixo e de valores de verdade estáticos, para um padrão lógico natural, contingente, movido pelo acaso, que permearia todos os processos evolutivos naturais.

Destacamos que, mesmo tendo influenciado o surgimento das lógicas não clássicas, a filosofia naturalista, que pretendia aproximar lógica e filosofia dos processos evolutivos, relacionando-as à vida, acabou sendo deixada de lado pelos avanços que se seguiriam no âmbito da Ciência Cognitiva. Acreditamos que isso tenha ocorrido principalmente devido ao que Dupuy chama de “fascínio pelo modelo”. Esse encanto pela modelagem algorítmica fez com que, após, a originária cibernética fosse superada, principalmente, a partir dos trabalhos de Herbert Simon, quando tivemos o início da Ciência Cognitiva conforme a concebemos hoje. Passou-se, então, a pressupor os processos cognitivos como reproduzíveis a partir de programação de funções normativas, sem a necessidade de se levar em consideração o plano biológico e corpóreo.

Em clara oposição às ideias relacionais da cibernética a abordagem simbólica, foca-se num produto final (resolução de um problema), ignorando por completo o processo. O processo cognitivo resume-se, portanto, ao processamento sequencial da informação de entrada e a consequente transformação desta informação numa saída, que é a

1 - A origem da “virada naturalista”, segundo Gonzalez e Broens (2011), pode ser datada na metade do séc. XIX, principalmente com a publicação da obra *A origem das espécies*, de Darwin. Conforme ressalta Dewey em seu artigo *The influence of Darwinism in Philosophy* (1909), Darwin teve grande influência no pensamento filosófico e científico, o que propiciou o surgimento de uma *nova lógica* de investigação, não mais pautada em pressupostos fixos, imutáveis, sobrenaturais e transcendentais. A virada naturalista trouxe novidades em filosofia como, por exemplo, a aproximação da reflexão filosófica dos problemas cotidianos (*os life problems*) e o diálogo com as ciências como a biologia, a química e a física em prol de uma concepção de mundo dinâmica, evolutiva e contingente.

resposta do sistema ou resolução do problema. (PINTO, 2007, p. 14)

Como vemos, a partir dos trabalhos de Marvin Minsky, Allen Newell e Herbert Simon na década de 1950, o principal intuito da Ciência Cognitiva passou a ser a simulação dos processos cognitivos em termos de meios e fins, manipulando tão somente regras e símbolos digitais. O chamado *General Problem Solver* (1957) foi um dos primeiros programas de computador concebidos para simular os processos inteligentes do pensamento humano em suas atividades de resolução de problemas. O modelo do computador digital seria, para Simon, a melhor imitação dos processos cognitivos e desse tipo de investigação, segundo ele, resultariam os futuros avanços na Ciência Cognitiva. A partir daí, inaugurou-se o que podemos chamar de paradigma simbólico/computacional da Ciência Cognitiva.

Esses avanços marcaram, conseqüentemente, o reconhecimento da Inteligência Artificial enquanto área de pesquisa. A inauguração da IA teria sido na reunião ocorrida no colégio de Darmouth, EUA, em 1956, na qual Newell e Simon apresentam o seu *Logic Theorist* e propõem a denominação de Inteligência Artificial (IA) para essa nova fase de trabalhos. Seguidamente, J. McCarthy e P. Hayes aperfeiçoam os trabalhos iniciais da promissora IA desenvolvendo o programa *Lisp*, em 1960, com pretensões de que esse fosse um sistema processador capaz de agir inteligentemente no mundo. (VIGNAUX, 1991)

Foi a partir do desenvolvimento da IA, nas últimas décadas, que toda a ideia de uma ciência da mente se desenvolveu. A IA proporcionou o passo fundamental para se tentar relacionar mentes e computadores e estabelecer o que passamos a chamar de “modelo computacional da mente”. Não fossem os desenvolvimentos e realizações da IA nas últimas décadas suas máquinas de jogar xadrez, demonstrar teoremas matemáticos, realizar diagnósticos médicos toda uma polêmica sobre a natureza da mente e da inteligência não teria surgido. (TEIXEIRA, 1998, p. 13)

Ainda segundo Teixeira (1998, P. 104), “a Inteligência Artificial simbólica herdou uma metafísica cartesiana ao estabelecer uma dualidade entre *software* e *hardware* como metáfora para conceber as relações mente/cérebro.” Portanto, a Ciência Cognitiva e a IA tomaram como ponto de partida metodológico o pressuposto da separação entre a mente e matéria, ou seja, tinham como princípio a ideia de que as capacidades cognitivas seriam instanciáveis, independentemente de sua base física/biológica. Assim, a mente seria uma funcionalidade passível de ser simulada por um processador central que manipulasse corretamente

algoritmos. Esse paradigma inspirou-se principalmente nas contribuições advindas de Alan Turing (1950) – Funcionalismo –, que permitiu teoricamente a concepção de uma máquina abstrata (Máquina de Turing) capaz de realizar qualquer tipo de operação manipulando apenas símbolos. Desse paradigma resultou o estabelecimento de duas posturas na IA e a consequente distinção entre “IA Forte” e “IA Fraca”: a versão “Forte” da IA postularia justamente a certeza de que a mente poderia ser integralmente reproduzida em um computador, incluindo estados conscientes, intencionais e comportamentos autônomos, pois o mental não passaria de um processo sequencial de procedimentos passíveis de transcrição algorítmica. A “IA Fraca”, por sua vez, defende uma postura menos radical. Resumidamente, podemos dizer que a versão “Fraca” da IA teria como principal pressuposto que as modelagens artificiais servem apenas para aprofundar e melhorar nosso conhecimento sobre a mente, mas sem ter o objetivo de criar inteligência autônoma equivalente à humana.

Após essa etapa inicial de trabalhos, um segundo momento marcou o desenvolvimento da Ciência Cognitiva e da IA. Foi o surgimento da abordagem chamada *conexionista*. Pretendendo-se colocar como alternativa à postura simbólico/funcionalista, o conexionismo baseou-se nos estudos de redes neuronais naturais para modelar os sistemas artificiais, criando, com isso, as *Redes Neurais Artificiais*. Em linhas gerais, podemos dizer que esse paradigma se opõe, em certa medida, à abordagem meramente sintático/simbólica das capacidades mentais, indicando para a necessidade de se considerar a emergência semântica que ocorre nos processos cognitivos. Enquanto que para a I.A. simbólica, pela simulação algorítmica (sem plausibilidade biológica) se chegaria ao conhecimento da mente, para as Redes Neurais Artificiais seria pela simulação do cérebro (também por meio de algoritmos, mas considerando a plausibilidade biológica) que teríamos uma adequada interpretação dos processos cognitivos.

A cognição, no viés conexionista, consistiria na emergência de estados globais em uma rede de componentes materiais simples. Esses elementos simples que podemos chamar de nódulos conectar-se-iam entre si sem a necessidade de uma central de gerenciamento responsável pelo processamento da informação. Assim, a mente seria resultado do funcionamento de toda essa rede constituída por esses nódulos.

A abordagem conexionista é uma tentativa de construir um modelo de mente mais próximo de sua realidade biológica. Embora estes sistemas não sejam um modelo completo do cérebro e de seu funcionamento, pode-se pelo menos dizer que eles são inspirados na estrutura do cérebro. (TEIXEIRA, 1998, p. 85).

Ao que parece, com o conexionismo tivemos uma aproximação entre a mente e a estrutura biológica, na medida em que se entende o processo de produção cognitiva como resultado estrito das conexões neuronais e da estrutura cerebral, rechaçando a postura da IA *simbólica* que preservava a múltipla instanciação da mente. Nesse sentido, para o conexionismo, o significado, enquanto resultado dos processos cognitivos, estaria intimamente ligado às ocorrências sinápticas do cérebro e não ao jogo simbólico, como pretendia a I.A. *simbólica*. No processo cognitivo, ao atingir o nível chamado de *subsímbolo*, o significado se descolaria do plano biológico e da estrutura cerebral, emergindo daí o conhecimento, sem, no entanto, deixar de estar relacionado com alguma parte específica na rede neural do cérebro. Apesar de o conexionismo ter se colocado como alternativa à I.A. *simbólica*, apontando para a importância cognitiva dos sistemas biológicos, seu pressuposto básico continua sendo o de que a formalização seria o melhor meio para se reproduzir a inteligência humana, sem considerar a corporeidade. Por isso, convém destacar que no conexionismo, mesmo considerando-se a plausibilidade biológica, ainda não temos a preocupação com as capacidades cognitivas típicas das interações ambientais e perceptuais ligadas ao plano da ação.

Em resumo, as abordagens tradicionais da Ciência Cognitiva e IA – *simbólica* e *conexionista* – não consideravam a relevância da dinâmica do plano da ação, da corporeidade e da constituição biológica. Nesse sentido, apresentaremos a proposta da *Cognição Incorporada e Situada (CIS)*, que surge na década de 1980, como alternativa à mencionada tradição, desenvolvendo uma teoria dinâmica dos processos cognitivos que admite interações entre corpo, cognição e mundo. (SMITH, 1999, p. 769).

Podemos dizer que a teoria da *Cognição Incorporada e Situada (CIS)* desenvolveu-se tendo como objetivo evitar as dificuldades enfrentadas pelo paradigma computacional/simbólico da IA, conforme destacaremos nos trabalhos de Andy Clark (1998, 2001, 2008), Thelen & Smith (1994), Tony Chemero (2007), Haselager & Gonzales (2006) e Port & Van Gelder (1995). Em linhas gerais, a proposta da CIS é de mostrar o lado dinâmico e contínuo dos processos cognitivos. A CIS, ao absorver contribuições de áreas como da filosofia, da psicologia do desenvolvimento, dos estudos da percepção, das neurociências e da robótica, constitui-se então como um programa de pesquisa interdisciplinar com o objetivo de desenvolver modelos inteligentes corpóreos, capazes de interagir com o ambiente e de resolver situações problema. Assim, o viés da *Cognição Incorporada e Situada* tem como principal pressuposto que os sistemas orgânicos são constituídos por capacidades interacionais com o ambiente, que não podem ser dissociadas de sua história evolutiva, de sua corporeidade e de seu contexto.

Andy Clark pode ser apontado como um dos principais proponentes do tipo de abor-

dagem situada e incorporada. O autor desenvolveu vários temas importantes sobre o papel epistêmico do plano da ação nas performances inteligentes. Dentre esses, destacamos a criação do conceito de *ação epistêmica* enquanto importante avanço no sentido de diluir as separações e divisões entre pensamento e ação. Ações epistêmicas seriam certos tipos de ações que se dirigem ao aperfeiçoamento de nossa *performance* mental: em outras palavras, são alterações que fazemos no mundo visando não adaptar-nos à necessidades impostas fisicamente pelo ambiente, mas visando facilitar o desempenho de nossas capacidades cognitivas. Como vemos, Clark assume uma postura continuísta, segundo a qual não haveria uma ruptura entre as capacidades cognitivas do plano da ação e as capacidades cognitivas do plano racional/teórico/formal. Para o autor as ações epistêmicas:

Alteram o mundo para ajudar e auxiliar os processos cognitivos enquanto reconhecimento e pesquisa. Meras ações pragmáticas, ao contrário, alteram o mundo porque alguma mudança física é desejável por si própria (por exemplo, a ação de tapar o buraco de uma represa). Ação epistêmica, sugerimos, exige a propagação de crédito epistêmico. Se, diante de uma tarefa, uma parte do mundo funciona como se fosse um processo feito na cabeça, não hesitaríamos em reconhecer como parte dos processos cognitivos, de forma que esta parte do mundo é (nós defendemos) parte dos processos cognitivos. Processos cognitivos não estão (todos) na cabeça! (CHALMERS & CLARK, 1998, p. 3-4)

Como vemos, Clark defende a ideia, segundo a qual os processos cognitivos não estão sediados em um órgão ou instância superior ou primaz. Tal ideia se torna fundamental para sua proposta de mente estendida, pois, para ele, haveria uma extensão inteligente da mente que extrapola os limites meramente cerebrais, racionais, *conscientes*, espalhando-se por todo o conjunto corpóreo de que somos constituídos. Apoiando-se nas investigações de Kirsh e Maglio (1994) sobre os padrões de comportamento de seres humanos ao jogar *Tetris*, Clark conclui que, nesse tipo de situação, temos presente uma coparticipação direta da ação do indivíduo em seu processo cognitivo.

Segundo Clark, a pesquisa elaborada por Kirsh e Maglio (CHALMERS & CLARK, 1998) nos remete ao comportamento de seres humanos em uma situação na qual estão jogando *Tetris*. Nessa pesquisa, tomou-se o pressuposto de que haveria dois tipos de conduta para solução do encaixe das peças. Uma primeira opção seria realizar a rotação mentalmente para, em seguida, executar a rotação física por meio de um botão de comando ligado à

tela de um computador. A outra opção de resolução seria realizar primeiro a rotação física para ver qual encaixe se adequa melhor ao contexto que se apresenta na tela. A conclusão à que Kirsh e Maglio chegam, e da qual Chalmers e Clark se apropriam, é a de que as rotações feitas com a utilização do botão de comando, ou seja, na prática, tem uma maior eficácia do que as realizadas mentalmente. Enquanto a rotação “prática” demora 300 milissegundos, a rotação “mental” demora 1000 milissegundos.

De acordo com o autor, a partir desse experimento, é possível percebermos a importância do plano da ação e do conhecimento prático para o aperfeiçoamento de performances cognitivas. Nesse sentido, parece ser adequada a defesa de que a mente não se reduz a um órgão central como o cérebro, por exemplo, mas que é própria da natureza da mente ser dinâmica e estendida ao corpo e ao mundo, ou seja, a mente seria algo encarnado, incorporado e situado que se desenvolveria sempre a partir da relação com o ambiente.

Também em favor da relevância cognitiva do plano da ação, destacamos o pensamento de Thelen & Smith (1994). As autoras se posicionam criticamente em relação às abordagens lineares e dualistas dos processos cognitivos, propondo uma perspectiva *sistêmica* no âmbito da psicologia do desenvolvimento, superando os paradigmas tradicionais ao considerar a relevância cognitiva da ação. Os estudos de Thelen & Smith debruçam-se principalmente sobre o comportamento humano, em especial, no desenvolvimento cognitivo de bebês. A abordagem sistêmica de Thelen & Smith busca compreender os processos cognitivos considerando o todo como um único sistema, constituído de sistemas, dentro de outros sistemas, sempre em relação dinâmica. Nesse sentido, tentam explicar o desenvolvimento enquanto um processo complexo, abrindo mão de pressupostos dualistas como, por exemplo, mente/corpo, função/estrutura, cérebro/comportamento, etc.

Na concepção das autoras, a atividade auto-organizada teria um papel fundamental no desenvolvimento das capacidades comportamentais motoras, perceptivas, cognitivas e sociais, pois estas enfatizam o caráter auto-organizado dos sistemas inteligentes justamente porque defendem uma postura segundo a qual a mente seria processo complexo, resultado da dinâmica histórica, o contextual, cultural, social, etc. Consequentemente, as capacidades orgânicas básicas são consideradas como tendo um papel fundamental na cognição, pois são os ajustes cognitivos e auto-organizativos primitivos que revelam a inteligência de nossa corporeidade.

Por exemplo, as nossas rótulas limitam os graus de movimento possível com as pernas, fazendo o equilíbrio e locomoção muito mais fácil. É apenas um pequeno exagero dizer que aprender a andar é fácil para os seres humanos, já que as pernas já

sabem como fazer. (THELEN & SMITH, 1994. Apud Chemero, 2007, p. 27)

Nessa abordagem, entende-se que a herança ontogenética está estreitamente ligada às capacidades cognitivas do conhecimento do senso comum, manifestadas no desenvolvimento de sistemas biológicos. Nesse sentido, ao analisarem o desenvolvimento de crianças, as autoras concluem, por exemplo, que a formação de categorias (a formação da capacidade de classificar coisas diferentes, mas que guardam certas características comuns) é uma capacidade cognitiva fundamental para operações mentais de qualquer tipo, forjadas ontogeneticamente, sem precedentes inatos.

Sem esta habilidade [de formação de categorias], não há base para operações mentais de qualquer espécie, tanto de pensamento, quanto de ação. A formação de categorias é o início primitivo da vida mental; a ontogenia da formação de categorias é a base para o desenvolvimento do comportamento. (THELEN & SMITH, 1994, p. 162)

Assim, vemos que a CIS tende a recuperar os processos biológicos nas investigações dos processos cognitivos e da mente. Thelen & Smith consideram os processos físicos os responsáveis por estimular e impulsionar o desenvolvimento, argumentando a favor de uma relação direta do sistema inteligente com o mundo, sem a necessidade estrita de representações que façam a mediação dos processos cognitivos. “Mentes não representam necessariamente o mundo, elas vivem nele e fazem parte da realidade física, uma realidade do eu incorporado” (THELEN & SMITH, 1994, p. 164). Com essa afirmação, vemos o quanto a abordagem sistêmica pode contribuir para uma reavaliação do papel do conhecimento comum nos processos cognitivos, pois parece ser plausível considerarmos estes como os primeiros tipos de manifestação cognitiva constatada em seres humanos. Como citamos no exemplo da formação de categorias em crianças, esse desenvolvimento se dá a partir da experiência com o mundo, sem a necessidade prévia de teorização ou de alguma faculdade inata, evidenciando o caráter incorporado e evolutivo da inteligência.

Em suas críticas às abordagens tradicionais e lineares dos processos de desenvolvimento, Thelen & Smith ainda argumentam que a cognição se trata de um processo dinâmico no qual não há sistemas que se sobressaem a outros ou que sejam centrais. Nesse viés, todos os componentes de um sistema teriam igual importância dentro do processo cognitivo, cada qual desempenhando seu papel adequadamente e de forma contínua e dinâmica. Segundo Thelen (2000, p. 8 – tradução nossa),

É precisamente essencial a continuidade dos sistemas dinâmicos incorporados e acoplados no tempo para fluir comportamento adaptativo que dá sentido à noção de cognição incorporada. Não há momento em que estes processos parem e outra coisa assuma. Assim, há boas razões para acreditar não somente nas origens sensorio-motoras da cognição, mas também na relação íntima e intrincada entre agir e pensar ao longo da vida.

Por meio da citação, podemos ver o quão relevante se torna o fator *tempo* no âmbito da perspectiva incorporada e situada. Segundo Port & Van Gelder (1995, p. 3 - tradução nossa),

O sistema cognitivo não é um computador, ele é um sistema dinâmico. Não se trata de algo no cérebro, isolado e encapsulado; ao contrário, ele é um todo composto por sistema nervoso, corpo e ambiente. O sistema cognitivo não é um manipular sequencial discreto de estruturas representacionais estáticas; ao contrário, é uma estrutura contingente de influência mútua e simultânea. O seu processo não figura no tempo arbitrário e discreto dos passos de um computador; ao contrário, ele manifesta-se no tempo real de mudança contínua no ambiente, no corpo e no sistema nervoso. O sistema cognitivo não interage com os outros aspectos do mundo pela passagem de mensagens ou comandos; ao contrário, está continuamente evoluindo com eles.

Podemos dizer que a proposta da CIS herdando contribuições da abordagem dinâmica assume a “mudança” como palavra-chave. A mudança se torna fundamental na medida em que os processos cognitivos passam a ser vistos como incorporados no tempo e em um contexto a partir dos quais se torna possível a evolução de um sistema. Em suma, a CIS aplicada à robótica pretende desenvolver agentes artificiais que evoluam e aprendam com o ambiente, adquirindo padrões de conduta ao lidarem de forma apropriada com desafios e imprevistos apresentados em situações não controladas.

A relevância cognitiva do corpo

De acordo com o exposto até o momento, a corporeidade passa a ser entendida como tendo um papel fundamental dentro dos processos cognitivos. Esse pressuposto se contrasta com toda a tradição dualista filosófica e o paradigma computacional/funcionalista da mente. Ao compreender uma continuidade entre corpo, cognição e ambiente, teríamos a vantagem de dissolver alguns problemas como, por exemplo, o da relação entre mente e corpo, tão caro à filosofia tradicional. No paradigma incorporado e situado, a cognição surgiria como *emergência* dos processos biológicos corporais, de forma contínua e natural. Quando dizemos que há uma relação entre corpo e cognição não queremos estabelecer uma identidade de propriedades (como, por exemplo, a tipo-tipo²). No viés incorporado e situado, corpo e cognição não seriam a mesma coisa, mas também não seriam coisas distintas e excludentes. Trata-se de compreender a cognição como um *continuum* auto-organizado, sempre em transformação pela dinâmica que a constitui como um todo.

De acordo com os trabalhos de Clark (1998), por exemplo, o corpo possuiria uma dinâmica própria, reveladora de sua flexibilidade e plasticidade, o que nos permite estabelecer distinções, por exemplo, entre o corpo de um robô e o corpo de um ser vivo. Os movimentos do corpo de um ser vivo possuem uma flexibilidade e plasticidade espontâneas que se adaptam dinamicamente em relação ao ambiente. Já os movimentos de um corpo robótico não teriam tal flexibilidade espontânea, pois necessitariam de uma série de regras e parâmetros para o reconhecimento do local, para poderem tentar adaptar-se. Tal flexibilidade e plasticidade espontânea do corpo de um ser vivo são expressas por Clark como sendo um comportamento *soft assembly*. Essa expressão pode ser definida como movimentos produzidos por nosso corpo, nos quais há uma mistura de sutileza e força ao mesmo tempo, que se dão de forma espontânea, sem previsão, digamos, consciente: trata-se de um tipo de conhecimento tácito do ambiente.

No plano da ação, cognição e corpo fundem-se em uma continuidade dinâmica da qual resultam performances inteligentes de alta eficiência em situações do senso comum que realizamos. Segundo Haselager & Gonzalez (2006, p. 9)

Se olharmos para o exemplo da ‘quebra do ovo’, o problema de como quebrar um ovo não se resume a uma consulta na base de dados de nosso conhecimento proposicional para formular um planejamento simples. Ao invés disso, o meio ambiente e suas pro-

2 - A Teoria da Identidade Tipo-tipo, criada por UT. Place e J.J Smart na década de 1950, afirma que cada tipo de estado mental corresponde a um tipo de estado cerebral. Esse viés materialista/fisicalista da mente propõe, portanto, uma redução dos estados mentais aos estados físicos, o que não ocorre na abordagem da Cognição Incorporada e Situada.

priedades possibilitam, selecionam ou explicitam certos comportamentos. A borda da vasilha permite a quebra da casca do ovo. O movimento que fazemos com nossos braços e mãos não é planejado e calculado, mas emerge da dinâmica intrínseca de nosso corpo, do peso do ovo, da influencia da gravidade e de muitos outros fatores. Nesse sentido, os aspectos gerais do organismo na situação do senso comum se auto-organizam em uma sequência apropriada de comportamentos emergentes.

Ressaltamos que as contribuições da CIS permitiram à IA e à robótica avanços significativos nos últimos anos, justamente por destacar o relevante papel do corpo nos processos cognitivos em nível de senso comum. Por isso, esse tipo de abordagem tem sido acolhida nas ultimas décadas, principalmente no âmbito da IA e da robótica, e já deixa de ser apenas uma abordagem alternativa, ganhando o *status* de teoria científica corrente. (GUDWIN, 2005, p.14).

Como vemos, a partir da abordagem da CIS, diminui-se a distância entre o comportamento de seres artificiais e o de seres vivos. Uma das principais questões que pretendemos colocar é: seria cabível atribuir moralidade e responsabilidade a seres artificiais? Uma das primeiras propostas para esse debate foi apresentada por Isaac Asimov, em sua obra de ficção científica, “*Eu, Robô*”, quando ainda não tínhamos tipos tão avançados de “humanoides” como os de hoje. Segundo as leis de Asimov: 1) um robô não pode ferir um ser humano ou, por omissão, permitir que um ser humano sofra algum mal; 2) um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que tais ordens contrariem a Primeira Lei; 3) um robô deve proteger sua própria existência, desde que tal proteção não entre em conflito com a Primeira e Segunda Leis. Atualmente, com os avanços na área da IA, as leis de Asimov parecem não dar conta deste que podemos chamar de dilema moral dos robôs, principalmente porque possuem um valor prescritivo ainda aliado a concepções tradicionais de moral e ética.

Com os atuais avanços da IA, é razoável pressupormos que robôs podem ser relativamente considerados autônomos de diversas maneiras, desde aqueles que executam tarefas que exigem alta precisão ou repetitividade como os utilizados em linha de produção, até aqueles que oferecem certo grau de risco à vida do homem, pois lidam com os seres humanos atuando como recepcionistas (robô *Reem*, 2011, da *Palm Robotics*), como interlocutores (robô *Zeno*, 2010, da *Hanson Robotics*) e como atores (robô *Geminoid F*, 2012, Universidade de Osaka), capazes de interagir com seres humanos. Nesse sentido, temos um problema que parece extrapolar os limites dos modelos éticos prescritivos, do *agir por dever*, conforme expressos pelo código criado por Asimov e pelas filosofias tradicionais.

Sugerimos que algumas filosofias contemporâneas podem trazer contribuições para o referido debate. Tentamos mostrar que, num viés filosófico pragmático, não dualista, é possível uma reflexão sobre a conduta no plano da ação não mais detida no domínio *racionalista/representacionista/internalista*. Entre essas filosofias contemporâneas, destacamos o pensamento de John Dewey e sua proposta pragmatista, que enfatiza a relevância do plano da experiência e da ação para a superação do dualismo tipo “mente *versus* corpo”, “eu *versus* ação”, na compreensão das questões morais e éticas.

O pragmatismo de John Dewey: uma teoria da ação *inusitada*

Ao tratar da moral e da ética, Dewey toma sempre como objeto de investigação a conduta dos indivíduos, enfatizando que ações e condutas não se devem unicamente a um *eu* interno do qual partem as coordenadas. Inicialmente, podemos dizer que um dos principais pressupostos de John Dewey é o de que a formação e a constituição da conduta dar-se-ia de forma contínua; além disso, o ambiente seria decisivo para direção e formação dos hábitos daquele que está atuando ou agindo. Assim, os estímulos externos (que incitam o agente a ter ações e reações) contribuem no sentido de auxiliar aquele que age a adquirir hábitos dentro do processo de aprendizado.

Segundo Dewey, as vertentes de tendência metafísica/internalista (como a kantiana, por exemplo, e as *éticas do dever*, em geral) possuem uma concepção moral fortemente dualista, pois não compreendem a formação moral em sua totalidade e continuidade entre agente e ambiente. O “eu” é fundamentalmente um “eu” corpóreo que age, um agente como um todo com o ambiente que o circunda. A proposta deweyana é de que, entre o “eu” e a ação, não há uma distinção do tipo meio e fim. Diferentemente, “há um arranjo circular. O ‘eu’ não é simples meio para produzir consequências, porque estas, quando de espécie moral, entram na formação do ‘eu’ e este na formação delas.” (Dewey, 1964, p. 134). Em suma, a constituição de uma identidade dependeria do processo de aprendizagem que envolve tanto o plano individual (“eu”), quanto o ambiental. Como vemos, a ambiência ganha um papel fundamental na discussão da moral e é apontada por Dewey como sendo também relevante para a constituição da *consciência*.

Está a região moral isolada do restante da atividade humana? Somente uma classe especial de objetivos e relações humanos tem valor moral? Essa conclusão é o resultado necessário da teoria [tradicional] segundo a qual nossa consciência e nosso

conhecimento morais são únicos na espécie. Mas se a consciência moral não é separada, então não se pode traçar qualquer linha rígida e firme na conduta isolando o reino da moral do reino não-moral. (DEWEY, 1964, p. 110)

A consciência, para Dewey, antes de ter um caráter subjetivo, interno, racional, possui um caráter objetivo que a constitui. Esse caráter objetivo da consciência é fundamental para que compreendamos o genuíno significado de um julgamento moral, pois, nas perspectivas filosóficas tradicionais, tal julgamento se dava em uma perspectiva *internalista*. Para Dewey, a consciência deveria ser compreendida, então, de forma objetiva, experiencial, o que permitiria à reflexão sobre a moralidade expandir seu alcance, não limitando o julgamento moral apenas a um procedimento punitivo.

Um ser humano é considerado responsável a fim de que possa aprender; a fim de que possa aprender não teórica e academicamente, mas de forma a modificar e – até certo ponto – refazer seu “eu” anterior [...] o que interessa é se é capaz de agir diferentemente na próxima vez, a importância prática e modificações no caráter humano é que torna importante a responsabilidade. (DEWEY, 1964, p. 153)

Em suma, podemos dizer que o autor pretende que todo julgamento moral considere aspectos como o contexto, a sociedade, a cultura, as redes relacionais, etc., que são cruciais para a constituição de uma determinada identidade. “A consciência genuína tem perspectiva objetiva; é atenção e cuidado inteligentes para com a qualidade de um ato, em vista de suas consequências para a felicidade geral; não é solicitude ansiosa para com o estado virtuoso do próprio indivíduo” (DEWEY, 1964, p. 120).

O dilema moral dos robôs à luz do pragmatismo deweyano

De acordo com o que expusemos até aqui, pretendemos agora mostrar que a proposta de John Dewey parece plausível de ser relacionada à conduta não somente humana. Nesse sentido, tentando compreender o âmbito da conduta de seres artificiais à luz das considerações deweyanas, pode-se dizer que, a partir do momento em que um robô, em dado momento, desempenha um padrão de ação como, por exemplo, o de andar de bicicleta, de

tocar um violino ou de recepcionar pessoas em um evento, ele o faz imitando um comportamento típico, que expressa determinado objetivo. Ressaltamos que tal objetivo provém estritamente de todo o processo interacional com contexto do qual o robô participa.

De acordo com o projeto da robótica atual, a saber, de criar robôs capazes de aprender e de alterar hábitos conforme a dinâmica ambiental, torna-se possível que os mesmos tenham uma conduta *responsável* e sejam *responsabilizados* por suas ações. Não estamos debatendo se robôs teriam ou não consciência, estamos apenas indicando que o julgamento moral não se deve exclusivamente a aspectos internos e subjetivos. Em outras palavras, da mesma forma que um homem pode ser caracterizado como responsável, um robô também poderia vir a sê-lo, se considerado um agente que assume determinada postura, em um determinado contexto, pois sua conduta implicaria consequências práticas. Nossa interpretação é a de que a filosofia de John Dewey parece não encontrar problemas nessa afirmação, pois a responsabilidade, tanto no âmbito humano, quanto no âmbito de um agente artificial, sempre seria dividida entre ambiente e agente.

O julgamento moral, nesse caso, não deveria servir meramente como instrumento para culpar aquele que errou: responsabilizar um indivíduo, como vimos, implicaria, antes de qualquer punição, a expectativa de mudança de conduta. Assim, quando um robô erra, responsabilizá-lo implicaria não meramente uma acusação ou condenação; indo além, a responsabilidade deve levar esse robô à necessidade de mudança de padrões de hábito. Nesse sentido, na medida em que robôs forem capazes de tomadas de decisão, de mudança de hábitos nesses moldes, podem ser considerados seres *responsáveis* por suas ações, mas concebendo sempre a *responsabilidade* diferentemente dos pressupostos tradicionais.

O que deve ser redefinida, ou talvez até banida, é a noção de imputação de culpa e penalização enquanto algo que recai sobre uma consciência prévia que planeja e executa uma ação. A concepção de uma mente interna que coordene as ações tidas como corpóreas é recusada por Dewey, em prol de uma definição de conhecimento e ação oriunda de um processo complexo que se inicia pelo fluxo contínuo da experiência. Em outras palavras, a sociedade e o contexto são também *responsáveis* pela *má ação cometida* por um indivíduo. Nesse sentido, a sociedade e o criador do robô também teriam um papel em sua atuação, por isso, uma análise moral em termos tradicionais não pode dar conta do complexo problema que envolve o dilema moral dos robôs. Trata de um processo complexo que envolve uma dinâmica relacional pautada no plano da ação, e não em instâncias e faculdades subjetivas responsáveis pela *agência* do robô. Nesse sentido, responsabilizar um robô implica, antes de tudo, compreender a totalidade da situação em que se deu o erro implicando que, ao invés de possuir um caráter punitivo, essa responsabilização possua um caráter educativo.

Considerações finais

Ao que parece, abordar o chamado “dilema moral dos robôs” implica uma renúncia a éticas dualistas, o que traz implicações morais no sentido de que conceitos como de responsabilidade e liberdade sejam tratados filosoficamente de maneira alternativa. Como vimos na filosofia de John Dewey, os procedimentos morais são um *continuum* dos procedimentos cognitivos, e a responsabilidade de um agente estaria ligada à sua capacidade de prever as consequências de seus próprios atos. O autor, em seu ideal democrático, sempre se mostrou avesso aos sistemas prisionais e punitivos, pois acreditava que um julgamento moral devesse ser compreendido como algo educativo e não meramente punitivo.

Se, por exemplo, um robô recepcionista como o *Reem* (2011) causasse mal a algum visitante desse evento, o que diríamos dessa situação? De alguma maneira, esse robô, um agente interacional que cometeu um erro, não teve previsão suficiente de sua ação. Talvez isso nos permita interpretá-lo como sendo, em algum sentido, *responsável* por sua ação, mas não com o intuito primordial de penalizá-lo, como é corrente em nossos códigos morais. Nesse sentido, a *responsabilidade moral* redefinida parece ser aplicável a robôs.

No entanto, é preciso que façamos uma ressalva em nossa análise, pois para Dewey a predisposição de natureza biológica é fundamental para o desenvolvimento de um padrão de conduta e o estabelecimento de uma *identidade pessoal*. Por isso, antes que cometamos alguma injustiça com o pragmatismo de Dewey, sublinhamos que, como ainda não temos *efetivamente* modelos biologicamente desenvolvidos, conceber os robôs como possuindo uma identidade pessoal, um agir consciente, ainda não é possível. Restringimo-nos apenas a relacionar Dewey e o dilema moral dos robôs nos aspectos mencionados aqui, segundo os quais a ação não é algo de origem estritamente interna, mas um processo dinâmico entre agente e ambiente.

Bibliografia

ASIMOV, I. *Eu, Robô*. Trad. Luiz Horácio da Mata. São Paulo: Circulo do Livro, 1976.

BROOKS, R. *Inteligência sem representação*. Tradução: Pedro Rocha de Oliveira e João de Fernandes Teixeira. Em: www.filosofiadamente.org/images/stories/textos/representacao.doc. Acesso em: 28/11/2011.

_____. Intelligence without representation. In: *Artificial Intelligence* 47 (1991), pp. 139-159.

CHEMERO, A. *Radical Embodied Cognitive Science*. Massachusetts: MIT/Bradford, 2009.

CLARK, A. *Being there: putting brain, body, and world together again*. Massachusetts: MIT/Bradford, 1998.

_____. *Mindware: An introduction to the philosophy of cognitive science*. New York: Oxford University Press, 2001.

_____. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press, 2008.

CLARK; A.; CHALMERS, D. The Extended Mind. In: *Analysis*, 58 (1), p.10-23, 1998.

DEWEY, J. *Democracia e educação: introdução à filosofia da educação*. 3. ed. Tradução Godofredo Rangel e Anísio Teixeira. São Paulo: Companhia Editora Nacional, 1959.

_____. *Experiência e natureza*. São Paulo: Abril Cultural, 1980.

_____. *Reconstrução em Filosofia*. São Paulo: Companhia Editora Nacional, 1959.

_____. *Human Nature and Conduct: an introduction to social psychology*. New York: Modern Library, 1950.

_____. *Teoria da vida moral*. São Paulo: IBRASA, 1964.

DUPUY, J.-P. *Nas origens das Ciências Cognitivas*. Tradução: Roberto Leal Ferreira. São Paulo: Editora Unesp, 1996.

GONZALEZ, M. E. Q.; BROENS, M. C. Darwin e a virada naturalista na filosofia. In: MORAES, J. Q. (Org.). *Darwin e a origem do homem*. Campinas: UNICAMP, 2011.

GUDWIN, R.R. *Novas Fronteiras na Inteligência Artificial e na Robótica*. In: *Dincom'2005*. Campinas. 2005. pp. 01-18. Disponível em: http://www.dca.fee.unicamp.br/~gudwin/ftp/publications/Dincom05_Gudwin.pdf. Acesso em: 20/09/2012

HASELAGER, W.F.G.; GONZALEZ, M.E.Q. Conhecimento comum e auto-organização. Disponível em: <http://www.nici.kun.nl/~haselag/publications/ConhecimentoComumAutoorg.pdf>. Acesso em: 10/07/2011

HASELAGER, W.F.G. Auto-organização e comportamento comum: opções e problemas. In: GONZALES, M.E.Q; D'OTTAVIANO, I.M.L; SOUZA, G.M. (orgs). *Auto-organização: estudos interdisciplinares*. Vol. 38. Campinas, Brasil: Coleção CLE, 2004. pp. 213-235.

HASELAGER, W.F.G. O mal estar do representacionismo: sete dores de cabeça da Ciência Cognitiva. In A. Ferreira, M.E.Q. Gonzalez & J.G. Coelho (Eds.). *Encontros com as Ciên-*

cias Cognitivas, Vol. 4. São Paulo: Coleção Estudos Cognitivos, 2004. pp. 105-120.

KIRSH, D.; MAGLIO, P. On distinguishing epistemic from pragmatic action. In: *Cognitive Science*, 18, pp. 513-549, University of California, San Diego, 1994. Disponível em <http://adrenaline.ucsd.edu/kirsh/publications.html>. Acesso em: 23/06/2012.

PINTO, S.M.C.R. A Natureza Histórica da Cognição: Debates filosóficos na Teoria dos Sistemas Dinâmicos na Ciência Cognitiva. Dissertação de Mestrado. Universidade do Porto. Portugal, 2007.

PORT, R.; VAN GELDER, T. *Mind as Motion: Explorations in the dynamics of cognition*. Massachusetts: MIT/Bradford, 1995.

SMITH, B.C. Situatedness/Embeddedness. In: WILSON, R.A.; FRANK, C.K. (orgs). *The MIT Encyclopedia of the Cognitive Sciences*. Massachusetts: MIT, 1999. pp. 769-770.

TEIXEIRA, A. A pedagogia de Dewey. In: *Vida e Educação*. São Paulo: Companhia Editora Nacional, 1959.

_____. Bases da teoria lógica de Dewey. *Revista Brasileira de Estudos Pedagógicos*. Rio de Janeiro, v. 23, n. 57, jan/mar, 1955 p. 3-27.

THELEN, E. Grounded in the world: developmental origins of the embodied mind. In: *Infancy*, 1 (1), 3-28, 2000.

THELEN, E.; SMITH, L. *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press, 1994.

VIGNAUX, G. *As Ciências Cognitivas: uma introdução*. Coleção Epistemologia e Sociedade. Tradução: Maria Manuela Guimarães. Lisboa: Instituto Piaget, 1991.