ULP Law Review | Revista de Direito da ULP
Vol. 14, n.º 1
doi: 10.46294/ulplr - rdulp.v14i1.7470

Moralidade Humana, Moralidade Maquinal: A *Programming Machine Ethics* Como Problema Jurídico

Ana Elisabete Ferreira[1]
Instituto Jurídico

I) O Substrato Biológico da Capacidade Valorativa Humana

i) A intersubjetividade humana é única por múltiplas razões, a primeira das quais puramente biológica[2]. "Os dados oferecidos à primeira pessoa que contribuem para a autonomia da consciência implicam referências constantes ao «eu», à história pessoal do sujeito, à perceção do seu corpo".[3]

O alcance desta afirmação, especulativamente complexa, evidencia-se nos esforços de programação artificial da decisão ética. A perspetiva de um *universo prostético*[4] testará os limiares das funções cognitivas humanas; o biónico superará as mais incríveis performances animais, os exoesqueletos darão lugar a autênticas exo-vivências que obliterarão quaisquer limitações endógenas, e o «deep learning» admitirá programar *softwares* realmente autónomos, totalmente independentes da nossa intervenção e decisão.

---

[1] Investigadora Colaboradora do Instituto Jurídico – Faculdade de Direito da Universidade de Coimbra (Portugal). Professora Adjunta Convidada do Instituto Politécnico de Leiria (Portugal). Responsável pela Secção de Direito Civil Médico do Instituto de Derecho Iberoamericano (Espanha). Especializada em Direito da Medicina e Pós-graduada em Direito da Farmácia e do Medicamento pelo Centro de Direito Biomédico (Portugal). Doutorada em Bioética pela Universidade Católica Portuguesa. Licenciada e Mestre em Direito pela Faculdade de Direito da Universidade de Coimbra.

[2] A. Damásio, *A Estranha Ordem das Coisas - A vida, os sentimentos e as culturas humanas,* (Lisboa: Temas e Debates, 2017), 45.

[3] J.P. Changeaux, *L´Homme de Vérité* (Paris: Odile Jacob, 2004), 121 – 123.

[4] M. Curado, *Luz Misteriosa – A Consciência no Mundo Físico* (Famalicão: Quasi Edições, 2007): 251 e ss.

O Presente promete, portanto, construir uma consciência artificial complexa, plataformas capazes de transformar estímulos exteriores em partículas integradas de vivência humana, transverter peças mecânicas em impulsos orgânicos, computar memórias – e depois, fatalmente, emoções – a partir de algoritmos, que tornarão objetos inanimados em brilhantes aprendizes com pensamento valorativo[5]. Ora, independentemente da objetividade dos resultados já obtidos nesta área, há um problema fundamental que, porventura, não tem sido suficientemente densificado: a programação ética das máquinas («programming machine ethics»)[6] parte do pressuposto de que a decisão eticamente adequada é culturalmente construída e *aprendida* – pelo que pode ser também *ensinada* (designadamente, às máquinas). Que concluir, porém, se este modo de olhar a moralidade estiver equivocado?

ii) A posição mais conhecida relativamente ao tema da origem do comportamento moral é aquela que nasceu como *o jardineiro* de Thomas Huxley[7] e veio desaguar no *gene egoísta* de Dawkins: numa palavra, a ideia de que o ser humano se torna ético por oposição à sua própria natureza[8], tendo de arrancar as *ervas daninhas* do seu *jardim* permanentemente, empenhando a sua vida política e social no sentido de se libertar do mundo darwiniano[9]. Esta conceção é compatível com a visão clássica do homem como uma «tábua rasa», que ao nascer possui apenas uma primeira natureza incompleta, a rematar com uma segunda natureza, a cultura. Nasce, pois, como uma tela em branco que a educação e aprendizagem irão *pintar* de uma determinada forma. Na síntese do eminente sociólogo Nikolas Rose, "their unspoken premise, for at least the past century has been that *human beings (…) come into the world unfinished* and that *our individual capacities, mores, values, thoughts, desires, emotions (in short, our mental lives), as well as our*

[5] E. Hildt, "Artificial Intelligence: Does Consciousness Matter?", *Frontiers in Psychology,* vol. 10 (2019): 1535.
[6] Luís Moniz Pereira, and Ari Saptawijaya, *Programming Machine Ethics* (Springer, 2016), 109.
[7] *Vide* De Waal, *Primates and Philosophers – How Morality Evolved,* Princeton University Press (2006*):* 7. Cfr. M. Patrão Neves, "Na Senda da Responsabilidade Moral", *in Poética do Mundo – Homenagem a Joaquim Cerqueira Gonçalves* (Dep. Fil. UL, org.) (Lisboa: Edições Colibri, 2001), 851 – 870.
[8] De Waal, F., cit., p. 9.
[9] Cfr. R. Dawkins, *A Devil´s Chaplain: Reflections on Hope, Lies, Science, and Love* (New York: Houghton Mifflin, 2003), 12 e ss.

group identities, family structures, loyalties to others, and so forth are shaped by upbringing, culture, society, and history"[10].

A cultura como força que conforma os seres humanos no sentido de superar a biologia é porventura ainda a doutrina dominante. Nas palavras de Francisco Ayala, "a distinctive human social organization is culture, which may be understood here as set of *non-strictly biological* human activities and creations. (…) *Culture (…) become the dominant mode of human evolution*"[11].

Trata-se da visão comum, que identifica a moralidade com uma construção cultural e consuetudinária, e supõe que *o bom* e *o bem* se aprendem e interpretam de um modo estritamente conjuntural. Steven Pinker denunciou-a como «o paradigma da *tábua rasa»,* segundo o qual "our ancestors (…) become moral by choice"[12].

iii) Há, porém, uma outra forma de ver a questão que, na senda da paleoantropologia contemporânea e com o decisivo contributo das neurociências, afirma a existência de uma moralidade inata, biologicamente ratificada.

Para compreendê-la é indispensável termos presente as lições de Konrad Lorenz sobre os obstáculos ao autoconhecimento dos seres humanos[13]: 1) que o primeiro obstáculo é literalmente *primitivo,* e se prende com a nossa ignorância sobre as nossas origens; 2) que o segundo obstáculo consiste na dificuldade em aceitar que o nosso comportamento obedece a leis de *causação natural* que são empiricamente verificáveis – e só a interpretação do *determinismo biológico,* e não a sua existência, é sindicável –; e que o terceiro obstáculo advém da herança da filosofia idealista, segundo a qual o mundo das

[10] N. Rose, and Abi-Rached, *Neuro – The New Brain Sciences and the Management of The Mind* (Princeton: Princeton University Press, 2013), p. 2 (realce nosso).
[11] F. J. Ayala, "The difference of being human: Morality" *in Proceedings of the National Academy of Sciences,* Vol. 107 (2010): 9015 – 9022.
[12] De Waal, *Primates and Philosophers,* cit., p. 6.
[13] Assim em K. Lorenz, *On Agression,* trad. Marjorie Wilson (London, Routledge, 2002), 214- 217. No mesmo sentido, M. Tomasello, *A Natural History of Human Thinking* (Harvard, Harvard University Press, 2014), 81 e ss.

leis naturais é um mundo sem valores, amoral, ou moralmente neutro, o que, manifestamente, não corresponde à realidade[14].

Os circuitos neuronais do cérebro regem-se e regem o corpo através de dois princípios fundamentais: a preservação e o bem-estar[15]. A preservação pressupõe absolutamente a cooperação, o cuidado próprio e o cuidado do outro. Os neurónios não fazem isto sozinhos, fazem-no com o auxílio e o bom conselho das hormonas e das proteínas, que são autênticos *agentes* do cérebro, operando transações, negociações, equilíbrios difíceis[16].

"Caring and caring for"[17] consubstanciam o substrato moral essencial do corpo, a profunda raiz da nossa moralidade. Todos os sistemas nervosos estão organizados no sentido de garantir a sobrevivência do corpo, de que são parte. Os animais que falham na preservação têm pouca *chance* de manter-se saudáveis e, assim, de passar os seus genes. O primeiro objetivo do corpo é, portanto, garantir a sua homeostase. Quando uma determinada necessidade é detetada, uma emoção sobrevém. A sede, por exemplo, é uma emoção primordial que pode gerar-se quando a corrente sanguínea apresenta um dado nível de sódio. É um sinal ou *marcador homestotático* de uma ameaça, que precisa ser sanada para não comprometer o regular funcionamento do organismo.

iv) O cuidado é provavelmente o aspeto mais complexo da moralidade. Como é que o cérebro mantém plataformas para os valores? Ou como o cérebro se importa com alguma coisa? A resposta começa na autopreservação. Como explica Churchland[18], cuidar é uma função pervasiva dos sistemas nervosos. Emoções primordiais nasceram para expressar necessidades primárias dos corpos, como fome ou frio. Tal como sucede como as pequenas células do nosso corpo, a sociabilidade humana desenvolve-se num contexto de empatia e justiça distributiva. Portanto, embora seja inegável o papel que os padrões morais eruditos e os sistemas sociais culturalmente construídos desempenham no comportamento humano, é importante lembrar que aquilo a que chamamos "comportamento ético" depende

[14] Changeux, and Ricoeur, *What Makes Us Think?* (Princeton: Princeton University Press, 2000), 217: "Let´s not call Kant – a pre-evolutionist philosopher – to the rescue in a discussion of evolution. (…) The «structuring» evolution of norms takes over epigenetically from the natural evolution of species".
[15] Churchland, *Braintrust,* cit., pp. 12 a 14.
[16] *Ibidem.*
[17] Churchland, *Braintrust,* cit, especialmente, pp. 27 a 62.
[18] *Ibidem.*

absolutamente do bom funcionamento de certas estruturas cognitivas e emocionais do cérebro.

Encontramo-nos aqui num campo da neuroética cujos estudos se concentram no surgimento da moralidade na mente humana. O facto de termos nascido numa atmosfera cooperativa, com centenas de milhares de anos de sedimentação, é um elemento-chave para entender o surgimento de um sentimento de pertença, subjacente ao que De Waal chamou de «community concern»[19].

Este modo de funcionamento é verificável também externamente, particularmente nos mamíferos, cuja produção ampliada de determinadas hormonas (como a ocitocina) veio determinar laços familiares e comunitários robustos, projetando o sentido do cuidado do eu para os indivíduos existencialmente mais próximos, e até a indivíduos de outros grupos e mesmo estranhos[20]. Com efeito, a produção de ocitocina confere ao animal certas emoções que o impelirão a cuidar do outro. Além de desempenhar um papel definitivo na génese do parto e na lactação, a produção de ocitocina incita a manter-se quente e seguro, além de ser responsável por uma emoção comummente designada como *ansiedade da separação*[21]. Em consequência, a proximidade relativamente ao outro deixa de ser uma opção, para passar a ser uma questão de sobrevivência[22].

Portanto, ao invés de termos desenvolvido a moral, *a partir do nada*, através da reflexão racional, recebemos uma enorme influência do nosso «plano de fundo» como animais sociais[23]. Embora a normatividade mais complexa dependa fortemente do seu contexto histórico, "partes de todas as fórmulas são universais"[24]. Somos, portanto, inatamente morais, o que significa que a nossa moralidade é natural.

---

[19] De Waal, *Primates And Philosophers – How Morality Evolved* (Princeton: Princeton University Press, 2006), 44-45.
[20] Churchland, *Braintrust,* pp. 30 e 31.
[21] *Idem,* p. 33.
[22] Megan Galbally, *et al.*, "The Role of Oxytocin in Mother-Infant Relations: A Systematic Review of Human Studies", *Harvard Review of Psychiatry*, Volume 19, Issue 1 (2011): 1 – 14.
[23] São palavras de em De Waal em De Waal, *The Bonobo and The Atheist – In Search of Humanism Among The Primates* (New York: Norton , and Company, 2013),13.
[24] Damásio, *A Estranha Ordem das Coisas,* cit., p. 47.

II. A construção de Padrões Morais de Conduta dos Seres Humanos

i) Não obstante o alegado acima, esta nossa *moralidade inata* não deve ser confundida com os nossos padrões morais de conduta construídos. O complexo moral humano apresenta um distintivo claro, a elaboração: "a move toward universal standards combined with an elaborate system of justification, monitoring, and punishment"[25]. Em suma, padrões. Padrões de compreensão e justificação, fiscalização comportamental, castigo e uma enorme sensibilidade à opinião dos outros. Os padrões morais têm sido vistos como a verdadeira moralidade. Como dizia Ernst Mayr, a ética *genuína* requer o fator cultural, a implementação da "pregação de um líder religioso ou de um filósofo", o pensamento dos "líderes culturais". Mayr concede que a tendência altruística é inata («inborn altruistic tendencies»), mas entende que só a cultura poderá operar um direcionamento correto dessas tendências – "redirecting (…) toward a new target: outsiders"[26].

Não obstante a dimensão culturalizante dos padrões aprendidos e repetidos, seria muito difícil explicar o nosso apelo pela justiça e pela equidade se não fossem as fortes reações emocionais do nosso corpo à sua falta[27].[28]

ii) Os inúmeros sentidos e distintos horizontes discursivos em que a moralidade é convocada sugerem uma análise complexa e uma parametrização difícil, quer do conceito, quer do seu objetivo circunstancial. Torna-se sinuoso convocar um sentido de moralidade que não tenha subjacente uma certa opção epistemológica (ou mesmo ideológica) e de natureza teorética previamente definida, sem simultaneamente a violentar com a adjetivação «artificial».

A história do pensamento crítico mostra-nos, quer um paradigma descritivo e cultural da moralidade, no seio do qual se observam códigos de conduta apresentados por certos

[25] De Waal, *The Bonobo and The Atheist,* cit., p. 14.
[26] E. Mayr, *What Evolution Is?* (London: Phoenix, 2002), 286.
[27] De Waal, *The Bonobo and The Atheist,* cit., p. 13.
[28] Cfr. P. Gilbert,"Evolution And Depression: Issues And Implications", *Psychological Medicine* (2006): 36, 287–297.

grupos ou sociedades[29], quer paradigmas normativos universalizantes[30], quer outros que se concentram na construção da humanidade do homem[31], através de processos mais ou menos evidentes de inclusão e exclusão de características individuais e de grupo, empiricamente verificáveis.

Os estudos contemporâneos, por sua vez, patenteiam uma relação estreita entre a moralidade e o comportamento[32] individual. O estudo da primeira desenvolve-se, muitas vezes, a partir de conceções psicológicas e/ou empírico-evolucionistas do comportamento. Aliás, um problema particular com o qual antropólogos e biólogos evolucionistas se depararam, no advento das suas disciplinas, foi o de saber como um comportamento civilizado pode emergir de um comportamento incivilizado e puramente *animal*[33]. Com efeito, só mais recentemente, a neuroética – contando com a tecnologia neurocientífica moderna que permitiu analisar o cérebro vivo, em funcionamento, em tempo real – viria a oferecer uma resposta acabada a esta questão. Esta resposta, não obstante, é de difícil encaixe no âmbito artificial, porque pressupõe o substrato evolucionário das células humanas[34].

Numa visão que parte das emoções primordiais para a formação de uma primeira consciência – aquela, precisamente, que irá permitir a assunção do pensamento valorativo[35] – torna-se especialmente difícil conjeturar que a mesma consciência valorativa pudesse assomar artificialmente. O tema tem sido amplamente debatido, mesmo considerando apenas o intervalo concreto dos últimos três lustros: em *Comment la matière devient conscience*[36], Edelman e Tononi, com base numa teoria antiga do primeiro autor,

---

[29] Por todos A. Honnet, and H. Joas, *Social Action and Human Nature,* trad., Raymond Meyer (Cambridge: Cambridge University Press, 1988), 90 e ss.

[30] *Vide* T. Aquinas, *Summa Theologica* (Irvine: XistPub., 2004), 83; A. Alves, "O Fundamento da Liberdade Humana em Santo Tomás de Aquino", *Synesis,* v. 3, n. 2 (2011): 1 – 18, 14.

[31] Assim em P. Sloterdijk, *Regras para o Parque Humano,* trad. Manuel Resende (Coimbra: Angelus Novus, 2007), 7 e ss., 51 e ss.

[32] B. Lindström, *et al.*, "The role of a «common is moral» heuristic in the stability and change of moral norms", *Journal of Experimental Psychology General* 147, no2 (2018): 228-242. L. Cosmides, , and J. Tooby, "From evolution to behavior: Evolutionary psychology as the missing link" in *The Latest on the Best: Essays on Evolution and Optimality,* John Dupre (ed.), MIT Press (1987): 276-306.

[33] J. Cartwright, *Evolution and Human Behavior: Darwinian Perspectives on Human Nature,* (GB: Palgrave, 2000), 16.

[34] *Vide* Dennett, D., *A Liberdade Evolui,* cit., pp. 70-77.

[35] J. W. Buckholtz , and R. Marois, "The Roots of Modern Justice: Cognitive and Neural Foundations of Social Norms and Their Enforcement", *Law and Neuroscience* (Jones, Schall, Shen, eds.), cit., pp. 525-588,

[36] G. Edelman, and G. Tononi, *Comment La Matière Devient Conscience* (Paris: Odile Jacob, 2000.

procuram explicar como a mente se torna consciente e, posto isso, o que é a consciência. Com efeito, o Nobel da Medicina dedicou boa parte da sua carreira a procurar demonstrar que a consciência depende da *capacidade de construir um cenário mental no presente*.

Esta capacidade "baseia-se na categorização percetiva dos influxos visuais e das outras informações sensoriais relativas ao mundo exterior"[37], não dependendo, num primeiro momento, que exista um sentimento de si ou uma linguagem construída. Na tese de Edelman, da qual também Denton e Damásio partem, é o trabalho dos órgãos sensoriais (progressivamente mais bem equipados) o que merece o louvor primeiro pela criação da consciência.

Na obra de Derek Denton, *As Emoções Primordiais*[38], são estas as responsáveis pela génese da consciência. As emoções primordiais geram *excitações imperiosas* e *intenções compulsivas* porque, quando uma emoção primordial se gera indica que o organismo está ameaçado, e precisa agir[39]. A sede, a dor ou a falta de ar são bons exemplos de emoções primordiais. Os neurónios têm a aptidão de detetar e vigiar a composição físico-química do sangue que aflui ao cérebro, porque a sua sobrevivência depende disso. Havendo alterações ao nível da composição sanguínea, é gerada uma excitação que impele a um dado comportamento. A sede, por exemplo, é essencialmente uma alteração da concentração de sódio no sangue, que impele o animal a procurar beber. Mas o ato de beber, quando acontece, não é um ato isolado e suficiente. Para matar a sede é necessário não apenas que o animal proceda à ingestão de água, mas também que tenha alguma *noção* disso. Especialmente, se pensarmos num contexto selvagem, em que a ingestão de água pode ser particularmente difícil e perigosa. O animal tem de saber que bebeu, que foi *ele mesmo* que bebeu, e que isso o faz sentir-se bem novamente.

Apoiado, nomeadamente, na obra do antropólogo Donald Griffin sobre as *Animal Minds,* Denton afirma que o animal *precisou de saber* quando estava a beber, ou a comer ou a fugir[40]. Esta é, portanto, a sua *consciência primária*.

[37] *Vide* Denton, D. cit., p. 22.
[38] *Idem,* pp. 49-74.
[39] *Idem,* p. 23.
[40] *Idem,* pp. 40-47.

Na obra de Damásio, a génese da consciência é provocada pelo ensejo homeostático, ou seja, a necessidade que o corpo tem de proteger a sua própria integridade.[41] "Outra implicação de os estados homeostáticos ideais serem a mais valiosa posse de um organismo vivo é o facto de a vantagem fundamental da consciência, a qualquer nível do fenómeno, derivar da melhoria da regulação da vida em ambientes cada vez mais complexos"[42]. Os seres vivos que, por força da necessidade de adaptação ao meio, foram desenvolvendo a consciência, evoluíram no sentido de procurar, não apenas a sua sobrevivência, mas um certo grau de bem-estar. Quanto mais aprimorada a consciência, mais rebuscado o conseguimento subjetivo da homeostase. Damásio é muito claro na afirmação de que a gestão fundamental da sobrevivência é "a causa primeira da emergência e evolução dos cérebros (…) e de tudo o que se seguiu ao desenvolvimento de cérebros"[43].

No que respeita à consciência alargada, por seu turno, há também amplo consenso entre os biólogos e neurocientistas mais proeminentes na atualidade, como Edelman, Changeux, Damásio ou Denton. Em primeiro lugar, a superação da teoria clássica de John Eccles, segundo a qual não existiria qualquer «eu», mas uma consciência autoconsciente. Em segundo lugar, a ideia de que a consciência alargada é uma espécie mais evoluída de consciência (relativamente à consciência primária ou primordial). Em terceiro lugar, a ideia de que a consciência alargada não é apenas um *upgrade* da primeira consciência, mas algo que detém a inteligência de se transcender autopoieticamente. *Se a consciência nuclear constitui o alicerce indispensável da consciência, a consciência alargada é o seu apogeu*[44].

iii) Recordemos que, em Damásio, a consciência exige um sentimento de si contemporâneo. A consciência nuclear – a primeira consciência – basta-se com um *modesto si,* que é um *fugaz sentimento do conhecer,* instantâneo e subtil. Na consciência alargada, o si acarreta consigo toda a bagagem biográfica, é o *si autobiográfico*. Por

[41] Damásio, A., *O Livro da Consciência,* especialmente, pp. 76-81.
[42] *Idem,* p. 82.
[43] *Idem,* p. 85.
[44] Damásio, *O Sentimento de Si,* cit., p. 227.

outras palavras, vem acompanhado da memória do passado, que pode ser reativada, e de uma certa ideia de pessoalidade. Inclui, também, aquilo que Changeux designou como «espaço de trabalho neuronal»[45] – a interligação e distribuição da atividade neuronal. A capacidade de aprender e criar registos e, seguidamente, a possibilidade de os reativar, proporcionam um plano consciente ampliado, inclusivamente capaz de usar as memórias passadas para criar imagens futuras[46]. Damásio sintetiza este processo na afirmação de que a consciência desenvolvida se forma a partir da consciência nuclear com duas estratégias: a primeira consiste em formar memórias como imagens mentais, a segunda abrange a manutenção ativa de numerosas imagens por um período de tempo substancial, de modo a que se mantenham disponíveis[47].

iv) Há, contudo, (pelo menos) uma hipótese alternativa não biológico-evolucionista para o problema da consciência. Na psicologia, o cognitivismo reage ao *behaviorismo*[48] [49] afirmando a necessidade de conhecer o pensamento em si mesmo, e não a partir do comportamento. O cognitivismo aparece frequentemente associado ao funcionalismo da engenharia de *software*. E não se trata de uma metáfora – para o cognitivismo computacional, o pensamento seria tudo o que o cérebro é, exceto a sua parte física (não obstante a parte não física *residir* na composição física).

Duas obras foram pioneiras ao explorar a hipótese de o funcionamento do cérebro humano se assemelhar ao de um computador: *A Logical Calculus of Ideas Immanent in Nervous Activity*, de Walter Pitts e Waren McCulloch[50] e *Computing Machinery And Intelligence (The Imitation Game)*, de Alan Turing[51]. Não é uma ideia *nova*, embora só se tenha aprimorado substancialmente a partir do final do século XX, com as investigações de

[45] Changeux, *L´Homme de Vérité*, cit., pp. 123 e ss.
[46] Damásio, *O Sentimento de Si*, cit., pp. 228 e 229.
[47] *Idem*, pp. 229-231.
[48] Sobre o behaviorismo, a obra magistral de B. Skinner, *Ciência e Comportamento Humano*, trad. João Carlos Todorov/Rodolfo Azzi (São Paulo: Martins Nunes, 11.ª edição, 2003).
[49] Para uma primeira aproximação, K. Candiotto, "Nova Síntese: um diálogo inacabado entre Pinker e Fodor", *Revista de Filosofia Aurora*, Curitiba, V. 22, n.º 30 (2010):153-177.
[50] *In Bulletin Of Mathematical Biophysics*, vol. 5 (1943): 115 – 133.
[51] *In Mind*, vol. 49 (1950): 433 – 460.

autores como Hilary Putnam, Ned Block e Jerry Fodor[52], e à medida a que foram sendo enjeitadas as teses *coneccionistas* clássicas, baseadas em analogias forçadas com o «computational networking» operado por máquinas digitais. Desde então, e apesar de toda a crítica que lhe é assacada, esta tese não cessou de ampliar-se e incrementar os seus argumentos.

Aquilo que podemos designar como «cognitivismo computacional», que opera uma analogia entre a relação entre o pensamento e o cérebro e a relação entre um computador e o seu *software*, não é particularmente popular entre os neurocientistas, nem entre os filósofos das neurociências, sobretudo depois da dura crítica operada de John Searle em *The Rediscovery of the Mind*[53].

Searle opera aí uma escalpelização filosófica das funções da sintaxe e da semântica que vai acompanhar toda a sua bibliografia até ao momento atual. Para o final da sua exposição, Searle deixa um considerando, à partida, muito perturbante – o de que o cérebro não processa informação. Quase como se nos dissesse que, ainda que pudéssemos objetar a todos os argumentos anteriores, este é *definitivo*. Trata-se, com efeito, de atacar o pilar principal do cognitivismo, precisamente, o de que o cérebro é um processador de informação.

Para atentar nesta hipótese alternativa há que ter todas as cautelas com este argumento. Searle impugna, em última instância, que o cérebro possa comportar-se como um computador digital, porque o cérebro *não é* um computador digital, e porque o computador digital *não é* um cérebro. No caso do computador, é sempre um agente externo a codificar as informações que posteriormente serão processadas; além disso, no cérebro todas as ideias têm uma carga simbólica essencial[54].

[52] A síntese em M. Rescorla, "The Computational Theory of Mind", in *Stanford Encyclopedia of Philosophy*, publicado primeiramente em outubro de 2015, disponível em http://plato.stanford.edu/entries/computational-mind/, acedido em 11-01-2016.
[53] J. R. Searle, *A Redescoberta da Mente*, trad. Eduardo Pereira e Ferreira (São Paulo: Martins Fontes, 1997), 306-318 (especialmente).
[54] Para uma introdução aos problemas essenciais do cognitivismo, por exemplo, R. Canal, "Quatro Objeções de John Searle ao Cognitivismo", *Kínesis*, Vol. I, n.º 1 (Março de 2009): 171 – 185; W. T. Fitch,"Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition", *Physics of Life Reviews*, 11 (2014): 329 – 364.

Contudo, em boa verdade, vários argumentos são aqui confundidos e as conclusões, porventura, falham o seu alvo: como sintetisa Rescorla, "Since classical computationalists need not claim (and usually do not claim) that the mind is a programmable general purpose computer, the objection is misdirected." [55]

Ou seja, que o cérebro não é um computador, e vice-versa, parece evidente. Tal não significa que não possam operar-se analogias entre os dois. (E analogias autênticas, não metáforas.) Os robalos não têm ouvidos, mas ouvem; as magnólias não têm sistema digestivo, mas alimentam-se e digerem; os catos não possuem sistema reprodutor nem mantêm relações sexuais, mas reproduzem-se.

Do mesmo modo que não se pode afirmar que um computador não pensa porque não detém cérebro, o facto de o cérebro não dispor de um processador digital idêntico ao de um computador não é razão suficiente para afirmar que o cérebro não processa informação. Portanto, subsiste a questão de fundo: há evidência de que o cérebro processe informação?

Em Steven Pinker, esta é uma hipótese não apenas exata, mas que verdadeiramente se impõe para ultrapassar a barreira óbvia das diferenças culturais das mentes. Tal só é possível, segundo o autor, se começarmos a pensar em mecanismos mentais universais, como *software* mental[56], em vez de procurarmos compreender a mente por recurso a comportamentos observados. Segundo Pinker, "something in the head must be capable of generating not just any combinations of words but highly systematic ones. That something is a kind of software, a generative grammar that can crank out new arrangements of words. (…) Once one starts to think about mental software instead of physical behavior, the radical differences among human cultures become far smaller (…) Universal mental mechanisms can underlie superficial variation across cultures."[57] De acordo com o autor, o facto de a moralidade *evidenciar que as categorias familiares de comportamento (…) certamente variam entre as culturas e precisam ser aprendidas,* não invalida que *os*

[55] M. Rescorla, "The Computational Theory of Mind" in *Stanford Encyclopedia of Philosophy*, publicado primeiramente em outubro de 2015, disponível em http://plato.stanford.edu/entries/computational-mind/, acedido em 11-01-2016.
[56] Pinker, S., *The Blank Slate*, pp. 33 a 37.
[57] *Idem*, pp. 36 e 37.

mecanismos mais profundos da computação mental que os geram podem ser universais e inatos[58].

Para o cognitivismo computacional, «a mente» é, afinal, uma «façon de parler»; designa o modo como olhamos para o todo da atividade cerebral, que, por ser aparentemente tão vasto, torna difícil acreditar que se possa concentrar num substrato físico único. Há muitos outros elementos mentais a sofrer o mesmo embate especulativo: memória, subjetividade, biografia. E é esse o ponto de retorno hermenêutico do fisicalismo – o de haver sempre algo que não *está lá* – e que remete para entidades não físicas, fatalmente ausentes dos nervos.

III. A Construção da Moralidade Artificial

i) Falar na construção de uma moralidade artificial é, de certo modo, colocar a questão *ao contrário* do cognitivismo computacional: já não se trata de saber se o cérebro é um computador, mas antes de saber se um computador é (poderá ser) um cérebro. A ética das máquinas teve um grande incremento com a automação de veículos, o avanço da tecnologia industrial e os dispositivos médicos. Saber quando intervir, quem privilegiar, quem sacrificar, qual é a medida aceitável na limitação da liberdade, da privacidade, da decisão humana, estão entre as questões mais controversas. Muitos poderão argumentar que as máquinas não são agentes éticos autênticos, mas essa é apenas uma discussão conceptual, na medida em que as máquinas estão a ser programadas, cada vez mais, para tomar decisões não mediadas por humanos e aí darão, efetivamente, um certo *sentido* às suas ações.

ii) Isso leva-nos ao problema de saber qual o padrão de princípios que fundamentará tal programação. O padrão do homem médio – baseado na resposta que a maioria dos humanos daria em cada situação[59] – é, por enquanto, o padrão em vigor, mas é um

[58] *Ibidem.*
[59] Nas palavras de Moniz Pereira: "by appropriate moral decisions we mean the ones that conform with those the majority of people make, based on empirical results". *Vide* Moniz Pereira , and Saptawijaya, *Programming Machine Ethics,* Springer International Publishing Switzerland AG (2016), 109.

padrão altamente contestado na ética como no direito. Os resultados empíricos já publicados ilustram o facto de que a média das decisões é puramente utilitarista: a maioria das pessoas pensa que é preferível sacrificar uma pessoa para salvar cinco. Obviamente, a maioria delas responde a essas perguntas sem informações objetivas e sem educação ética essencial.

Um juízo sensitivo, porém, implica uma ponderação, porque todas as alternativas podem ser, em concreto, antiéticas, e porque, muitas vezes, a melhor opção no plano ético continua a ser antijurídica.

Experiências como a da «máquina moral» do MIT[60] parecem programar um autómato com base na decisão da maioria, mas, na verdade, temos apenas uma decisão do programador por essa opção utilitária, já que as alternativas foram, todas elas, previamente dadas e não adaptáveis.

iii) Ora, se não for possível mimetizar a experiência ética humana – pelo menos enquanto *softwares* não detiverem sistemas nervoso e endócrino – existe uma solução melhor para a programação?

Julgamos que sim. Tal opção deve privilegiar uma arquitetura "humana", não ensinando aos robôs ações corretas e erradas individualmente, porque não há evidências de que os padrões situacionais tendam a repetir-se e é impossível prever todos os dilemas morais que uma máquina enfrentará em cada contexto. Um «stop» automático obrigatório (ou um outro mecanismo de «kill switch») numa situação de conflito pode parecer uma resposta interessante, mas deixa de o ser quando se trata de um comboio autónomo e atrás vem outro comboio. A construção humana de padrões morais é complexa, claro, mas baseia-se sempre em princípios normativos, portanto, deve ser esse o ponto de partida na programação. Para ser ética, a programação deve ser onto-antropológica:

α) Densificando princípios bioéticos – como o respeito pela autonomia humana, a beneficência, a não maleficência e a justiça;

[60] https://www.moralmachine.net/

β) Respeitando os princípios jurídicos centrais – os princípios da legalidade, da liberdade, da dignidade humana, do estado de direito democrático, da operacionalidade dos direitos fundamentais;

γ) Considerando a integridade comunitária – a compaixão, o cuidado mútuo, o laço social, a solidariedade e a proteção especial para as pessoas vulneráveis.

IV) A *Programming Machine Ethics* como problema especificamente jurídico: por uma programação baseada em Princípios Normativos

A mais recente Resolução do Parlamento Europeu, de 20 de outubro de 2020, que contém recomendações à Comissão sobre o regime relativo aos aspetos éticos da inteligência artificial, da robótica e das tecnologias conexas (2020/2012(INL) toma importantes opções na definição conceptual de problemas complexos para o Direito, como é o caso do que se entende por «autonomia» quando referido a um robô[61]. Neste contexto, é implementado um conjunto de regras com um objetivo definido *a priori*: «uma inteligência artificial antropocêntrica e antropogénica».[62] Como compreender e tornar efetivo este objetivo?

Uma ideia fica absolutamente clara: é necessário que o humano nunca perca o controlo e a supervisão da tecnologia artificial. Portanto, voltamos à questão que indicámos como mais importante: 1) é fundamental o respeito pela autonomia humana, pela beneficência, pela não maleficência e a justiça: deve sempre salvaguardar-se o direito de recusa em ser tratado por um robô, o direito a uma programação baseada na evitação do dano (a montante, com boa programação ética e a jusante com sistemas efetivos de ressarcimento de danos), e o estrito respeito pela equidade no acesso, em termos análogos ao acesso a cuidados de saúde e medicamentos, por exemplo.

[61] Sobre esta problemática já nos pronunciámos em A. E. Ferreira, "Problemas Ético-Jurídicos Concernentes aos Danos Causados por Robôs Autónomos na Saúde" *in Cadernos Ibero-Americanos de Direito Sanitário* (São Paulo, 2020): 12-24 e em A. E. Ferreira , and A. D. Pereira, "Uma Ética para a Medicina Pós-humana – Propostas ético-jurídicas para a mediação das relações entre humanos e robôs na saúde" *in Responsabilidade Civil e Medicina*, Rosenvald, Menezes, Dadalto (eds.), (São Paulo: Editora Foco, 2020), 1-20.

[62] https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_PT.html#title1

Tal não é, ainda, suficiente. Muitas atuações robóticas poderão ser eticamente aceitáveis e, porém, ilícitas e não justificadas juridicamente (e vice-versa). A programação e a utilização da inteligência artificial têm de respeitar os princípios da legalidade, da dignidade humana, do Estado de Direito democrático e da aplicabilidade direta dos direitos fundamentais. Tal far-se-á, em primeiro lugar, através da definição de um sistema de registo claro dos autómatos mais avançados (devemos impreterivelmente saber, com transparência, o que está a ser feito no domínio da inteligência artificial). Para tal seria fundamental constituir-se uma Agência Internacional (em termos análogos à Agência Europeia do Medicamento[63]), a quem coubesse concretamente, receber e analisar propostas de colocação de autómatos no mercado, analisando os respetivos ensaios e fiscalizando a sua eficácia e a sua segurança, com poderes de supervisão, de veto e de retirada do mercado.

Com efeito, um dos aspetos mais marcantes da «nova robótica» – *rectius*, aquela que lança mão de expedientes de inteligência artificial avançada – é a sua potencialidade danosa, particularmente no domínio dos «riscos desconhecidos» e mesmo «inconhecíveis»[64]. Ora, o tratamento destes danos tem de ser feito em duas frentes: a montante, através de uma programação ética e jurídica rigorosa; a jusante, através de um sistema de ressarcimento de danos independente de culpa, pois é evidente que a tradicional doutrina da «culpa in vigilando» (responsabilidade pelo obrigado à vigilância de coisas ou equipamentos) será, no caso dos autómatos mais avançados, insuficiente[65].

---

[63] "A EMA tem como principais responsabilidades autorizar e controlar medicamentos na UE. As empresas solicitam uma autorização única de introdução no mercado, que é emitida pela Comissão Europeia. Se obtiverem essa autorização, as empresas podem comercializar o medicamento em questão em todos os países da UE e do EEE." *Vide* https://europa.eu/european-union/about-eu/agencies/ema_pt . A ideia de criar uma Agência Europeia para a Robótica não é uma ideia nova (vide https://www.europarl.europa.eu/news/pt/press-room/20170210IPR61808/eurodeputados-querem-regras-europeias-sobre-robos-e-inteligencia-artificial?quizBaseUrl=https%3A%2F%2Fquizweb.europarl.europa.eu), porém, continua no plano das intenções.

[64] M. A., Hogg, "Liability of Unknown Risks: A Common Law Perspective" in 15th Annual Conference on Tort Law, Viena, ECTIL – European Centre of Tort and Insurance Law (2016): 26 a 28.

[65] Designadamente, em Portugal: no nosso contexto jurídico, quem tiver em seu poder coisa móvel ou imóvel, com o dever de a vigiar, responde pelos danos que a coisa ou os animais causarem, mas esta responsabilidade é afastada se provar que nenhuma culpa houve da sua parte (v.g., que não pôde intervir na decisão do autómato) ou que os danos se teriam igualmente produzido ainda que não houvesse culpa sua. Considerar a utilização de robôs uma «atividade perigosa» poderá também não ser o caminho, pois a responsabilidade civil será afastada se o utilizador mostrar que empregou todas as providências exigidas pelas circunstâncias com o fim de os prevenir (programação adequada, uso adequado, cumprimento dos deveres de *updating*, etc). *Vide* o nosso A. E. Ferreira, "Responsabilidade civil extracontratual por danos

Preparar uma arquitetura ético-jurídica capaz de garantir 1) a máxima evitação dos danos e 2) o ótimo ressarcimento dos danos é fundamental, mas é ainda exíguo, porque um olhar atento sobre a potencialidade danosa da utilização de autómatos inteligentes não pode limitar-se à perspetiva jurídica dos danos e às circunstâncias (limitadas) da relevância jurídica «tout court» das consequências desta utilização. É necessário considerar a preservação do laço social e da integridade comunitária como valores intrínsecos ao modo de vida dos humanos.

A compaixão, o cuidado mútuo, o laço social, a solidariedade e a proteção especial para as pessoas vulneráveis são valores transversais nas comunidades modernas de «sapiens», que o convívio com robôs não tem o direito de violar. Um futuro em que a prestação de cuidados de saúde ou de educação, por exemplo, fique desprovida de laços humanos é um futuro em que a narrativa humana, no seu sentido ôntico essencial, estará ameaçada. Esta questão de planeamento, aparentemente lateral, é absolutamente indispensável para que prossigamos juntos, de modo intersubjetivo, num desiderato comunitário e solidário, que o Direito ele mesmo se tem esforçado por potenciar, intergeracionalmente.

É provável que este objetivo só possa alcançar-se com um progresso significativo nos modos de programação. Uma programação de modo meramente lógico ou silogístico não conseguirá dotar as máquinas de capacidade de compreensão do que aqui está em causa. É difícil afirmar que já estejamos neste caminho. Há já exigentes trabalhos de programação que se afastam da lógica tradicional, em favor de uma programação dita «difusa» ou «multivalorada», e que permitem aos *softwares* uma resposta diferente daquela que imediatamente descende das premissas, assim criando um sistema de «múltiplas verdades» ou verdades difusas/nebulosas, através do afastamento do modo tradicional «se-se, logo».[66]

causados por robôs autónomos: breves reflexões", *Revista Portuguesa do Dano Corporal,* Coimbra, a.25 n.º 27 (2016): 39-63, e mais recentemente, com idêntica análise e conclusões, M. M. Barbosa, "O futuro da responsabilidade civil desafiada pela inteligência artificial: as dificuldades dos modelos tradicionais e caminhos de solução", *Revista de Direito Civil V 2* (2020): 261-306 (sobretudo pp, 273 e ss).

[66] Por todos, Julían-Iranzo *et al.*, "A Fuzzy Logic Programming Environment for Managing Similarity and Truth Degrees" *in XIV Jornadas sobre Programación y Lenguajes* (S. ESCOBAR, ed.), EPTCS 173 (2015): 71–86.

Mas é certo que, com enorme probabilidade, a programação baseada em princípios normativos essenciais levará, finalmente, ao *empoderamento* das máquinas. Por empoderamento[67] entende-se a formalização de uma motivação intrínseca, isto é, a promoção da liberdade de se aproximar dos princípios[68]. Isso fará com que as máquinas desejem executar ações, não apenas silogisticamente, mas *da melhor maneira possível,* de acordo com os princípios, instigando a sua compreensão.

O empoderamento das máquinas torná-las-á mais humanas, sem contradição: quando estimulamos a "iniciativa do robô", fazemo-lo através de princípios genéricos o suficiente para que o *software* seja capaz de aplicá-los em situações novas[69]. Assim, espera-se que o robô não apenas seja capaz de responder de acordo com situações predefinidas, mas também que seja capaz de gerar novos objetivos e diretivas conforme necessário em novas situações[70]. Em segundo lugar, uma programação baseada em princípios normativos terá a finalidade específica de quebrar a equivalência de ações: como deve um robô escolher de entre várias possibilidades de ação diferentes quando todas produzem essencialmente o mesmo resultado desejado, apenas de maneiras diferentes? Só poderá fazê-lo a partir de uma noção clara de hierarquia normativa.

Finalmente, espera-se que o empoderamento seja testado em zonas seguras para esse efeito, isto é, que a possibilidade de ir gerando e testando alternativas para compreender a mais adequada seja efetuada, não apenas «in silico», mas «in vivo», em espaços de teste com segurança e adequados para o efeito, um pouco como vem acontecendo com os veículos que se autoconduzem. Como os robôs apresentam, geralmente, uma realidade preceptiva distinta da dos seres humanos, a construção de uma linguagem comum permanece um desafio: não podemos pretender ingenuamente que um software compreenda, por exemplo, o conceito de dano e as várias formas de o infligir a partir de

[67] Salge, and Polani, "Empowerment As Replacement for the Three Laws of Robotics" in Frontiers in Robotics and AI, vol 4, art 25 (jun. 2017): 1 – 16.
[68] *Ibidem*: "While elsewhere we studied involved single-agent scenarios in detail, here, we present proof-of-principle scenarios demonstrating how empowerment interpreted in light of these perspectives allows one to specify core concepts with a similar aim as Asimov's Three Laws of Robotics in an operational way. Importantly, this route does not depend on having to establish an explicit verbalized understanding of human language and conventions in the robots. Also, it incorporates the ability to take into account a rich variety of different situations and types of robotic embodiment."
[69] *Vide* E. Topol, *Deep Medicine – How Artificial Intelligence Can Make Healthcare Human Again* (New York: Basic Books 2019), 268.
[70] Salge, and Polani, "Empowerment…", cit, p. 4.

um conjunto limitado de dados, se o robô não puder testá-lo e aprendê-lo empiricamente, em zonas de teste real. Ensaiar alternativas até que seja o próprio *software* a designar a que lhe parece mais segura e eficaz, justificando essa escolha através de mecanismos de rastreabilidade programados, é indispensável para que deixemos os calabouços do silogismo simples. (Afinal, é assim que nós humanos evoluímos, também.)

BIBLIOGRAFIA:

Andersen, Michael, and Andersen, Susan Leigh. "Robot be good - Autonomous machines will soon play a big role in our lives. It's time they learned how to behave ethically." *Scientific American,* (October 2010).

Andersen, Michael, and Andersen, Susan Leigh. *Machine Ethics*. New York: Cambridge University Press, 2011.

Armstrong, Stuart, and Sotala, Kaj. "How We´re Predicting AI – Or Failing To" *in Beyond AI: Artificial Dreams,* edited by Jan Romportl, Pavel Ircing, Eva Zackova, Michal Polak, Radek Schuster. Pilsen: University of West Bohemia, 2012.

Aroso Linhares, José Manuel. "Is Law's Practical-Cultural Project Condemned To Fail The Test Of 'Contextual Congruence'? A Dialogue With Hans Albert´s Social Engineering". *In Towards Recognition of Minority groups – Legal and Communication Strategies,* edited by Marek Zirk-Sadowsky *et al.*, 210 – 220.

Aroso Linhares, José Manuel. "Jurisdição, diferendo e «área aberta»: A caminho de uma "Teoria" do direito como moldura?" *In Estudos em Homenagem ao Prof. Doutor Jorge de Figueiredo Dias,* coordenados por Manuel da Costa Andrade, Maria João Antunes e Susana Aires de Sousa. Vol. 4 (2009): 443 – 477.

Baertschi, Bernard. "Human Dignity as a Component of a Long-Lasting and Widespread Conceptual Construct." *Bioethical Inquiry,* Springer (22 April 2014): 1–11. http://www.unige.ch/medecine/ieh2/files/5314/3472/9172/Dignity.pdf.

Baertschi. *La Neuroéthique: Ce Que Les Neurosciences Font A Nous Conceptions Morales*. Paris: Éditions La Découverte, 2009.

Barbosa, Mafalda Miranda. "O futuro da responsabilidade civil desafiada pela inteligência artificial: as dificuldades dos modelos tradicionais e caminhos de solução." *Revista de Direito Civil,* V. 2 (2020): 261-306.

Bartz, Jennifer A., Zaki, Jamil, Bolger, Niall, and Ochsner, Kevin N. "Social effects of oxytocin in humans: context and person matter." *Trends in Cognitive Sciences,* vol. 15, no 7 (Jul. 2011): 301 – 310.

Bar-Yosef, Ofer. "The Upper Paleolithic Revolution." *Annual Review of Anthropology,* vol. 31 (2002): 363 – 393.

Batey, Robert. "Minority Report and the Law of Attempt." *Ohio State Journal Of Criminal Law,* vol. 1 (2004): 689 – 698.

Beck, Ulrich. *Sociedade de Risco Mundial – em busca da segurança perdida*. Translated by Marian Toldy e Teresa Toldy. Lisboa: Edições 70, 2015.

Belzung, Catherine. *Biologia das Emoções*. Translated by Armando Pereira da Silva. Lisboa, Instituto Piaget, 2010.

Ben-Menahem, Hanina, and Ben-Menahem, Yemima. "Law and Science – Reflections." *Science in Context,* 12 (1999): 227 — 243.

Bensoussan, Alain, and Bensoussan, Jérémy, *Droit des robots*. Bruxelles: Larcier, 2015.

Bergstrom, Theodore C. "Evolution of Social Behavior: Individual and Group Selection." *Journal of Economic Perspectives,* Volume 16, Number 2, Spring (2002): 67 – 88.

Boehm, Christopher. *Moral Origins – The Evolution of Virtue, Altruism and Shame*. New York: Basic Books, 2012.

Boladeras, Margarida. "Introducción."In *Hans Albert, Razón Crítica y Prática Social*. Barcelona: Ediciones Paidós, 2002.

Bostrom, *Nick , and* Yudkowsky, E. "The Ethics Of Artificial Intelligence." *In Cambridge Handbook of Artificial Intelligence* (eds. William Ramsey , and Keith Frankish). Cambridge: Cambridge University Press, 2011).

Bostrom, Nick , and Sandberg, Anders. "Cognitive Enhancement: Methods, Ethics, Regulatory Challenges." *Science and Engineering Ethics,* vol. 15 (2009): 311–341.

Brass, Marcel, and Haggard, Patrick. "To Do or Not to Do: The Neural Signature of Self-Control." *The Journal of Neuroscience,* 27 (34), (2007): 9141 – 9145.

Changeaux, Jean-Pierre. "Sinaptic Epigenesis and the Evolution of Higher Brain Functions." *In Epigenetics, Brain and Behavior* (Paolo Sassone-Corsi, Yves Christen, eds.). London: Springer, 2012.

Changeaux, Jean-Pierre. *L´Homme de Vérité.* Paris: Odile Jacob, 2004.

Changeux, Jean-Pierre, *L´Homme Neuronal*. Paris: Pluriel, 2012.

Changeux, Jean-Pierre , and Ricoeur, Paul. *What Makes Us Think?* Princeton: Princeton University Press, 2000.

Chitolina, Claudinei Luiz, Pereira, José Aparecido, *Pinto,* Rodrigo Hayasi (orgs.). *Mente, Cérebro e Consciência – Um confronto entre Filosofia e Ciência,* Prefácio. São Paulo: Paco Editorial, 2015.

Chomsky, Noam. *Language and Mind* (3th edition). Oxford, Oxford University Press, 2006.

Chomsky, Noam. *On Nature and Language* (Adriana Belleti, Luigi Rizzi, eds.). Oxford: Oxford University Press, 2002.

Churchland, Patricia S. *Braintrust – What Neuroscience Tells Us About Morality.* Princeton: Princeton University Press, 2011.

Churchland, Patricia Smith. "Moral decision-making and the brain." *In Neuroethics: Defining the Issues in Theory, Practice, and Policy* (Judy Illes, ed.). Oxford: Oxford University Press, 2006.

Churchland, Patricia Smith. *Neurophilosophy – Toward a Unified Science of the Mind/Brain.* Massachssets: MIT Press, 1986.

Churchland, Patricia. "The Big Questions: Do We Have Free Will?" *In Law and Neuroscience* (Jones, Schall, Shen, eds.). New York, Wolters Kluwer, 2014.

Dahiyat, Emad Abdel Rahim. "Intelligent agents and liability: is it a doctrinal problem or merely a problem of explanation?" *Artificial Intelligence and Law*, nr. 18 (2010): 103 – 121.

Dalgleish, Tim. "The Emotional Brain." *Nature Reviews Neuroscience,* Vol 5 (July 2004): 582 – 589.

Damásio, António. *A Estranha Ordem das Coisas: a Vida, os Sentimentos e as Culturas Humanas.* Lisboa: Temas e Debates, 2017.

Damásio, António. "Neuroscience and Ethics: intersections." *The American Journal of Bioethics,* no 7 (2007): 3 – 7.

Damásio, António. *O Livro da Consciência: a construção do Cérebro Consciente*. Translated by Luís Oliveira Santos. Lisboa: Temas e Debates, 2010.

Damásio, António. *O Sentimento de Si: o Corpo, a Emoção e a Neurobiologia da Consciência*. Lisboa: Publicações Europa-América, 2008.

Damasio, H, T. Grabowski, R. Frank, A. M. Galaburda and A. R. Damasio. "The return of Phineas Gage: clues about the brain from the skull of a famous patient." Science 264 (1994): 1102-1105.

Dancy, Jonathan. *Ethics Without Principles*. Oxford: Oxford University Press, 2006.

Davidson, Donald. "Actions, Reasons and Causes." *Essays on Actions and Events: Philosophical Essays*. Oxford: Oxford University Press, 2001.

Devillers, Laurence. *Des Robots et des Hommes*. Paris: Plon, 2017.

De Waal, Frans, *The Bonobo and The Atheist – In Search of Humanism Among The Primates*. New York: Norton , and Company, 2013.

De Waal, Frans. *Primates and Philosophers – How Morality Evolved*. Princeton: Princeton University Press, 2006.

Dennett, Daniel. "Quinning Qualia." *In Consciousness in Modern Science,* edited by A. Marcel and E. Bisiach. Oxford: Oxford University Press, 1988.

Dennett, Daniel. *A Liberdade Evolui,* traduzido por Jorge Beleza. Lisboa: Temas e Debates, 2005.

Dennett, Daniel. *Darwin´s Dangerous Idea. Evolutions and the Meanings of Life*. London: Penguin Books, 1995.

Denton, Derek. *As Emoções Primordiais: A Emergência da Consciência*. Traduzido por Rui Pacheco. Lisboa: Instituto Piaget, 2010.

Derrida, Jacques. *Força de Lei*. Traduzido por Fernanda Bernardo. Porto: Campo das Letras, 2003.

Harcourt, Bernard. E. E*xposed: Desire and Disobedience in the Digital Age*. Harvard: Harvard University Press, 2015.

Hart, H. L. A. "Positivism and the Separation of Law and Morals." *Harvard Law Review,* vol. 71, no4 (**1958)**: 593 – 629.

Hart, Herbert L. A. *O Conceito de Direito*. Traduzido por A. Ribeiro Mendes. Lisboa: Fundação Calouste Gulbenkian, 2011.

Honneth, Axel. "Decentered Autonomy: The Subject After the Fall". *In Disrespect: The Normative Foundations of Critical Theory,* traduzido para Inglês por John Farrell, 261 – 271. Cambridge: Polity Press, 2007. [Original publicado como "Dezentrierte Autonomie. Moralphilosophische Konsequenzen aus der modernen Subjektkritik" *in Zur Verteidigung der Vernunft gegen ihre Liebhaber und Verächter* (Christoph Menke, Martin Seel, eds.). Frankfurt/Mein: Suhrkamp, 1993.]

Honneth, Axel. *Pathologies of Reason: On the Legacy of Critical Theory*. Columbia: Columbia University Press, 2009.

Honneth, Axel. *The I in We – Studies in the Theory of Recognition*. Translated by Joseph Ganahl. Cambridge: Polity Press, 2012.

Julían-Iranzo *et al.* "A Fuzzy Logic Programming Environment for Managing Similarity and Truth Degrees." *In XIV Jornadas sobre Programación y Lenguajes,* editado por S. Escobar, 71–86. EPTCS 173, 2015.

Lorenz, Konrad. *On Agression*. Traduzido para o Inglês por Marjorie Wilson. London: Routledge, 2002.

Loureiro, João Carlos. "Bios, Tempo(s) e Mundo(s): algumas reflexões sobre valores, interesses e riscos no campo biomédico." *In As Novas Questões em Torno da Vida e da Morte em Direito Penal. Uma Perspectiva Integrada,* organizado por José de Faria Costa e Inês Fernandes Godinho, 195 – 230. Coimbra: Wolters Kluwer/Coimbra Editora, 2010.

Loureiro, João Carlos. "Da sociedade técnica de massas à sociedade de risco: prevenção, precaução e tecnociência. Algumas questões juspublicísticas." *Estudos em Homenagem ao Prof. Doutor Rogério Soares, Studia Iuridica* 61, (2001): 797 – 891.

Loureiro, João Carlos. "Pessoa e Doença Mental." *Boletim da Faculdade de Direito* (Universidade de Coimbra), vol. 81 (2005):145 – 187.

Loureiro, João Carlos. "Prometeu, Golem & Companhia: Bioconstituição e Corporeidade Numa *Sociedade* (Mundial) De Risco." *Boletim Da Faculdade De Direito,* Universidade de Coimbra, vol. 85 (2009): 151 – 196.

Lüderssen, Klaus. "Wir Konnen Nicht Anders." *Frankfurt Allgemeine Zeitung,* 04-11-2003. http://www.faz.net/aktuell/feuilleton/hirnforschung-wir-koennen-nicht-anders-1134281-p3.html.

Luhmann, Niklas. *A Sociological Theory of Law*. Traduzido por Elizabeth King-Utz and Martin Albrow. New York: Routledge, 2014.

McCauley, Lee. "The Frankenstein Complex and Asimov's Three Laws." *Association for the Advancement of Artificial Intelligence* (online review), (2007): 9–14, https://www.aaai.org/Papers/Workshops/2007/WS-07-07/WS07-07-003.pdf .

McDermott, Drew. "Why Ethics is a High Hurdle for AI." *In North American Conference on Computers and Philosophy (NA-CAP)*. Bloomington, Indiana. July, 2008, 1–8. http://www.cs.yale.edu/homes/dvm/papers/ethical-machine.pdf.

Moll, Jorge, Zahn, Roland, Oliveira-Souza, Ricardo de, Krueger, Frank, and Grafman, Jordan. "The neural basis of human moral cognition." *Nature Reviews Neuroscience*, no 6 (October 2005): 799-809.

Moniz Pereira, Lenaerts, Martinez-Vaquero. "Guilt Emotion Enhances Cooperation in Evolving Multi-Agent Systems." (2016): 8 (não publicado – o texto foi-nos cedido pelo primeiro autor).

Moniz Pereira, Luís, and Saptawijaya, Ari. *Programming Machine Ethics*. Springer: International Publishing Switzerland AG, 2016.

Moniz Pereira, Luís. *A Máquina Iluminada, Cognição e Computação*. Porto: Fronteira do Caos, 2016.

Pagallo, Ugo. *The Laws of Robots: Crimes, Contracts, and Torts*. Springer, 2013.

Pardo, Michael S., and Patterson, Dennis. *Minds, Brains and Law – The Conceptual Foundations on Law and Neuroscience*. Oxford: Oxford University Press, 2013.

Pinker, Steven. *The Blank Slate – the modern denial of human nature*. London: Penguin Books, 2002.

Reis Marques, Mário. "A dignidade humana como prius axiomático." *Estudos em Homenagem ao Prof. Doutor Jorge de Figueiredo Dias,* organizados por Manuel da Costa Andrade, Maria João Antunes, Susana Aires de Sousa, vol. IV, Studia Iuridica 101 (2010): 541– 566.

Rescorla, Michael. "The Computational Theory of Mind." *In Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/computational-mind/.

Rodotà, Stefano. *Il Diritto Di Avere Diritti*. Roma-Bari: Laterza, 2012.

Rodotà, Stefano. *La Vita e le Regole – Tra Diritto e Non Diritto*. Milano: Feltrinelli Editore, 2006.

Salge, and Polani. "Empowerment As Replacement for the Three Laws of Robotics." *Frontiers in Robotics and AI,* vol 4, art 25, (jun. 2017): 1 – 16.

Timmermans, Stefan, and Epstein, Steven. "A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization." *Annual Review of Sociology,* Vol. 36 (Aug. 2010): 69 – 89.

Tomasello, Michael. "Why Don't Apes Point?" *Roots of Human Sociality: Culture, Cognition and Interaction*. Edited by N. J. Enfield, and S. C. Levinson, Oxford, and New York: Berg, 2006.

Tomasello, Michael. *A Natural History of Human Thinking*. Harvard: Harvard University Press, 2014.

Tomasello, Michael, Carpenter, Malinda, and Liszkowski, Ulf. "A New Look at Infant Pointing." *Child Development,* Volume 78, Number 3 (May/June 2007): 705 – 722.

Tononi, Giulio. *Comment la matière devient conscience*. Paris: Odile Jacob, 2000.

Toth, Amy L., Robinson, Gene E. "Evo-Devo and The Evolution Of Social Behavior." *Trends in Genetics,* Vol.23, No.7 (2007): 334 – 341.