

**REVISTA BRASILEIRA DE POLÍTICAS PÚBLICAS**  
**BRAZILIAN JOURNAL OF PUBLIC POLICY**

**Modelos de moralidade**  
Models of morality

Molly J. Crockett

**VOLUME 8 • Nº 2 • AGO • 2018**  
**DOSSIÊ ESPECIAL: INDUÇÃO DE COMPORTAMENTOS**  
**(NEUROLAW): DIREITO, PSICOLOGIA E NEUROCIÊNCIA**

# Sumário

<b>EDITORIAL</b> .....	24
<b>O Direito na fronteira da razão: Psicologia, neurociência e economia comportamental</b> .....	24
Patrícia Perrone Campos Mello e Sergio Nojiri	
<b>I. NEURODIREITO: COGNIÇÃO, EMOÇÃO, JUÍZOS MORAIS E CIÊNCIA</b> .....	26
<b>PENSAR DIREITO E EMOÇÃO: UMA CARTOGRAFIA</b> .....	28
Nevita Maria Pessoa de Aquino Franca Luna	
<b>NEURODIREITO: O INÍCIO, O FIM E O MEIO</b> .....	49
Carlos Marden e Leonardo Martins Wykrota	
<b>ENSAIO JURÍDICO SOBRE A RACIONALIDADE HUMANA: MAIORES, CAPAZES E IRRACIONAIS</b> .....	65
André Perin Schmidt Neto e Eugênio Facchini Neto	
<b>DIVERGÊNCIAS DE PRINCÍPIO: ARGUMENTOS JURÍDICOS E MORAIS EM UM CENÁRIO DE     DESACORDOS SOCIAIS</b> .....	90
André Matos de Almeida Oliveira, Pâmela de Rezende Côrtes e Leonardo Martins Wykrota	
<b>CONSILIÊNCIA E A POSSIBILIDADE DO NEURODIREITO: DA DESCONFIANÇA À RECONCILIAÇÃO     DISCIPLINAR</b> .....	117
Thaís de Bessa Gontijo de Oliveira e Renato César Cardoso	
<b>MODELOS DE MORALIDADE</b> .....	144
Molly J. Crockett	
<b>A INFELIZ BUSCA POR FELICIDADE NO DIREITO</b> .....	154
Úrsula Simões da Costa Cunha Vasconcellost, Noel Struchiner e Ivar Hannikainen	
<b>ALÉM DA LIBERDADE: PERSPECTIVAS EM NIETZSCHE</b> .....	178
Lucas Costa de Oliveira	
<b>A MEDIAÇÃO DE CONFLITOS SOB A PERSPECTIVA DO DESENVOLVIMENTO HUMANO: AS     CONTRIBUIÇÕES DA PSICOLOGIA POSITIVA</b> .....	193
Simone de Biazzi Ávila Batista da Silveira e Deise Brião Ferraz	
<b>NEUROIMAGIOLOGIA E AVALIAÇÃO DE RESPONSABILIDADE</b> .....	213
Nicole A. Vincent	

<b>II. NUDGES: INDUÇÃO DE COMPORTAMENTOS E POLÍTICAS PÚBLICAS .....</b>	<b>233</b>
<b>ANÁLISE CRÍTICA DA ORIENTAÇÃO DE CIDADÃOS COMO MÉTODO PARA OTIMIZAR DECISÕES PÚBLICAS POR MEIO DA TÉCNICA NUDGE.....</b>	<b>235</b>
Luciana Cristina Souza, Karen Tobias França Ramos e Sônia Carolina Romão Viana Perdigão	
<b>POLÍTICAS PÚBLICAS E O DEVER DE MONITORAMENTO: “LEVANDO OS DIREITOS A SÉRIO” .....</b>	<b>252</b>
Ana Paula de Barcellos	
<b>NUDGES E POLÍTICAS PÚBLICAS: UM MECANISMO DE COMBATE AO TRABALHO EM CONDIÇÃO ANÁLOGA À DE ESCRAVO .....</b>	<b>267</b>
Amanda Carolina Souza Silva, Débhora Renata Nunes Rodrigues e Saul Duarte Tibaldi	
<b>REDUZINDO A TRIBUTAÇÃO COGNITIVA: LIÇÕES COMPORTAMENTAIS PARA A DIMINUIÇÃO DOS EFEITOS PSICOLÓGICOS ADVERSOS DA POBREZA.....</b>	<b>288</b>
Leandro Novais e Silva, Luiz Felipe Drummond Teixeira, Gabriel Salgueiro Soares e Otávio Augusto Andrade Santos	
<b>POLÍTICAS PÚBLICAS EM SUICÍDIO: DO PATERNALISMO CLÁSSICO AO PATERNALISMO LIBERTÁRIO E NUDGING .....</b>	<b>327</b>
Davi de Paiva Costa Tangerino, Gabriel Cabral e Henrique Olive	
<b>NUDGES COMO POLÍTICA PÚBLICA PARA AUMENTAR O ESCASSO NÚMERO DE DOADORES DE ÓRGÃOS PARA TRANSPLANTE .....</b>	<b>369</b>
Roberta Marina Cioatto e Adriana de Alencar Gomes Pinheiro	
<b>OS PROGRAMAS DE INTEGRIDADE PARA CONTRATAÇÃO COM A ADMINISTRAÇÃO PÚBLICA ESTADUAL: NUDGE OU OBRIGAÇÃO LEGAL? UM OLHAR SOBRE AS DUAS PERSPECTIVAS .....</b>	<b>386</b>
Cíntia Muniz Rebouças de Alencar Araripe e Raquel Cavalcanti Ramos Machado	
<b>PATERNALISMO LIBERTÁRIO E PROTEÇÃO JURÍDICA DO AMBIENTE: POR QUE PROTEGER O AMBIENTE TAMBÉM DEVE SER PROTEGER AS LIBERDADES? .....</b>	<b>406</b>
Mariana Carvalho Victor Coelho e Patryck de Araujo Ayala	
<b>POLÍTICAS PÚBLICAS BASEADAS EM EVIDÊNCIAS COMPORTAMENTAIS: REFLEXÕES A PARTIR DO PROJETO DE LEI 488/2017 DO SENADO .....</b>	<b>429</b>
Pâmela de Rezende Côrtes, André Matos de Almeida Oliveira e Fabiano Teodoro de Rezende Lara	
<b>III. ECONOMIA COMPORTAMENTAL: VIESES COGNITIVOS E POLÍTICAS PÚBLICAS .....</b>	<b>455</b>
<b>ECONOMIA COMPORTAMENTAL E DIREITO: A RACIONALIDADE EM MUDANÇA .....</b>	<b>457</b>
Marcia Carla Pereira Ribeiro e Victor Hugo Domingues	
<b>VIESES COGNITIVOS E DESENHO DE POLÍTICAS PÚBLICAS.....</b>	<b>473</b>
Benjamin Miranda Tabak e Pedro Henrique Rincon Amaral	

<b>A NEUROCIÊNCIA DA MORALIDADE NA TOMADA DE DECISÕES JURÍDICAS COMPLEXAS E NO DESENHO DE POLÍTICAS PÚBLICAS .....</b>	<b>493</b>
Erik Navarro Wolkart	
<b>DESVIO DE CARÁTER OU SIMPLEMENTE HUMANO? ECONOMIA COMPORTAMENTAL APLICADA AO COMPORTAMENTO DESONESTO.....</b>	<b>524</b>
Diana Orghian, Gabriel Cabral, André Pinto e Alessandra Fontana	
<b>POLÍTICAS PÚBLICAS E A CONCRETIZAÇÃO DE DIREITOS SOCIAIS: TOMADA DE DECISÃO, ARQUITETURA DE ESCOLHAS E EFETIVIDADE .....</b>	<b>543</b>
Ana Elizabeth Neirão Reymão e Ricardo dos Santos Caçapietra	
<b>BEHAVIORAL ECONOMICS E DIREITO DO CONSUMIDOR: NOVAS PERSPECTIVAS PARA O ENFRENTAMENTO DO SUPERENDIVIDAMENTO .....</b>	<b>568</b>
Samir Alves Daura	
<b>A EDUCAÇÃO FORMAL PARA O CONSUMO É GARANTIA PARA UMA PRESENÇA REFLETIDA DO CONSUMIDOR NO MERCADO? UMA ANÁLISE COM BASE NA BEHAVIORAL LAW AND ECONOMICS (ECONOMIA COMPORTAMENTAL) .....</b>	<b>600</b>
Marcia Carla Pereira Ribeiro e Edson Mitsuo Tiujo	
<b>LIBET, DETERMINISMO E CONSUMO: AS INFLUÊNCIAS DO MARKETING E A RELEVÂNCIA DA DELIBERAÇÃO CONSCIENTE NA SUPERAÇÃO CONDICIONAL DE HÁBITOS DE CONSUMO PERIGOSOS.....</b>	<b>616</b>
Émilien Vilas Boas Reis e Leonardo Cordeiro de Gusmão	
<b>CIÊNCIA DO DIREITO TRIBUTÁRIO, ECONOMIA COMPORTAMENTAL E EXTRAFISCALIDADE.....</b>	<b>640</b>
Hugo de Brito Machado Segundo	
<b>IV. COMPORTAMENTO JUDICIAL: INFLUÊNCIA DE FATORES EXTRAJURÍDICOS .....</b>	<b>660</b>
<b>FATORES METAPROCESSUAIS E SUAS INFLUÊNCIAS PARA A FORMAÇÃO DA DECISÃO JUDICIAL .....</b>	<b>662</b>
Rogério Roberto Gonçalves de Abreu, Lúcio Grassi de Gouveia e Virgínia Colares	
<b>“A VIDA COMO ELA É”: COMPORTAMENTO ESTRATÉGICO NAS CORTES.....</b>	<b>689</b>
Patrícia Perrone Campos Mello	
<b>A COMPOSIÇÃO DO ÓRGÃO COLEGIADO E SEUS EFEITOS NA TOMADA DE DECISÃO .....</b>	<b>720</b>
André Garcia Leão Reis Valadares	
<b>DAS 11 ILHAS AO CENTRO DO ARQUIPÉLAGO: OS SUPERPODERES DO PRESIDENTE DO STF DURANTE O RECESSO JUDICIAL E FÉRIAS .....</b>	<b>741</b>
José Mário Wanderley Gomes Neto e Flávia Danielle Santiago Lima	

<b>RAZÃO, EMOÇÃO E DELIBERAÇÃO: AS ADEQUAÇÕES REGIMENTAIS DO SUPERIOR TRIBUNAL DE JUSTIÇA PARA A FORMAÇÃO DE PRECEDENTES EFICAZES .....</b>	<b>758</b>
Peter Panutto e Lana Olivi Chaim	
<b>HEURÍSTICA DE ANCORAGEM E FIXAÇÃO DE DANOS MORAIS EM JUIZADOS ESPECIAIS CÍVEIS NO RIO DE JANEIRO: UMA NOVA ANÁLISE .....</b>	<b>778</b>
Fernando Leal e Leandro Molhano Ribeiro	
<b>LA PROTECCIÓN DE LOS DERECHOS POLÍTICOS FRENTE A LAS FUNCIONES DISCIPLINARIAS DE LAS AUTORIDADES ADMINISTRATIVAS: SUBSIDIARIEDAD Y DEFERENCIA EN EL SISTEMA INTERAMERICANO DE DERECHOS HUMANOS.....</b>	<b>801</b>
Jorge Ernesto Roa Roa	
<b>V. A INFLUÊNCIA DO GÊNERO NO PROCESSO DECISÓRIO JUDICIAL.....</b>	<b>824</b>
<b>COMO OS JUÍZES DECIDEM OS CASOS DE ESTUPRO? ANÁLISANDO SENTENÇAS SOB A PERSPECTIVA DE VIESES E ESTEREÓTIPOS DE GÊNERO .....</b>	<b>826</b>
Gabriela Perissinotto de Almeida e Sérgio Nojiri	
<b>GÊNERO E COMPORTAMENTO JUDICIAL NO SUPREMO TRIBUNAL FEDERAL: OS MINISTROS CONFIAM MENOS EM RELATORAS MULHERES?.....</b>	<b>855</b>
Juliana Cesario Alvim Gomes, Rafaela Nogueira e Diego Werneck Arguelhes	
<b>HÉRCULES, HERMES E A PEQUENA SEREIA: UMA REFLEXÃO SOBRE ESTEREÓTIPOS DE GÊNERO, SUBPRESENTAÇÃO DAS MULHERES NOS TRIBUNAIS E (I)LEGITIMIDADE DEMOCRÁTICA DO PODER JUDICIÁRIO.....</b>	<b>878</b>
Jane Reis Gonçalves Pereira e Renan Medeiros de Oliveira	
<b>PRISÃO CAUTELAR DE GESTANTES: ANÁLISE DO FUNDAMENTO FILOSÓFICO DA DECISÃO DO HABEAS CORPUS N. 143.641 .....</b>	<b>912</b>
Artur César Souza e Giovania Tatibana de Souza	
<b>VI. NEURODIREITO APLICADO AO DIREITO E AO PROCESSO PENAL.....</b>	<b>926</b>
<b>CÉREBROS QUE PUNEM: UMA REVISÃO CRÍTICA DA NEUROCIÊNCIA DA PUNIÇÃO .....</b>	<b>928</b>
Ricardo de Lins e Horta	
<b>A INTUIÇÃO DO DOLO EM DIREITO PENAL: CORRELATOS NEURAI DA TEORIA DA MENTE, RACIOCÍNIO INDUTIVO E A GARANTIA DA CONVICÇÃO JUSTIFICADA.....</b>	<b>946</b>
Thiago Dias de Matos Diniz e Renato César Cardoso	
<b>AS COMUNIDADES EPISTÊMICAS PENAI E A PRODUÇÃO LEGISLATIVA EM MATÉRIA CRIMINAL.....</b>	<b>961</b>
Stéphane Enguéléguélé	

<b>DELINQUÊNCIA JUVENIL: RELAÇÕES ENTRE DESENVOLVIMENTO, FUNÇÕES EXECUTIVAS E COMPORTAMENTO SOCIAL NA ADOLESCÊNCIA .....</b>	<b>980</b>
André Vilela Komatsu, Rafaelle CS Costa e Marina Rezende Bazon	
<b>LÍMITES TEMPORALES A LAS PENAS PRIVATIVAS DE LIBERTAD ATENDIENDO AL DESARROLLO PSICOSOCIAL.....</b>	<b>1001</b>
Silvio Cuneo Nash	
<b>NEUROLAW E AS PERSPECTIVAS PARA UMA ANÁLISE OBJETIVA DO COMPORTAMENTO SUGESTIONADO: REPERCUSSÃO DAS FALSAS MEMÓRIAS NA ESFERA PENAL.....</b>	<b>1017</b>
Mariana Dionísio de Andrade, Marina Andrade Cartaxo e Rafael Gonçalves Mota	
<b>A FALIBILIDADE DA MEMÓRIA NOS RELATOS TESTEMUNHAIS AS IMPLICAÇÕES DAS FALSAS MEMÓRIAS NO CONTEXTO DOS CRIMES CONTRA A DIGNIDADE SEXUAL .....</b>	<b>1036</b>
Caroline Navas Viana	
<b>A (IR)REPETIBILIDADE DA PROVA PENAL DEPENDENTE DA MEMÓRIA: UMA DISCUSSÃO COM BASE NA PSICOLOGIA DO TESTEMUNHO.....</b>	<b>1058</b>
William Weber Ceconello, Gustavo Noronha de Avila e Lilian Milnitsky Stein	

**RESUMO**

Os dilemas morais geram conflitos entre duas tradições: a consequencialista, que avalia ações baseadas em seus resultados, e a deontológica, que avalia as ações em si. Elas se assemelham de forma marcante a duas distintas estruturas de decisão: um sistema baseado em modelos que seleciona ações fundadas em inferências acerca de suas consequências; e um sistema livre de modelos que seleciona ações com base em seu histórico de incentivos. Aqui, considero como esses sistemas, juntamente com um sistema pavloviano que responde reflexivamente a recompensas e punições, podem iluminar enigmas na psicologia moral.

**Palavras-chave:** Moralidade; Dilemas morais; Ética consequencialista e ética deontológica; Sistema pavloviano; Tomada de decisão.

**ABSTRACT**

Moral dilemmas engender conflicts between two traditions: consequentialism, which evaluates actions based on their outcomes, and deontology, which evaluates actions themselves. These strikingly resemble two distinct decision-making architectures: a model-based system that selects actions based on inferences about their consequences; and a model-free system that selects actions based on their reinforcement history. Here, I consider how these systems, along with a Pavlovian system that responds reflexively to rewards and punishments, can illuminate puzzles in moral psychology.

**Keywords:** Morality; Moral dilemmas; consequentialism ethic and deontology ethic; Pavlovian system; decision making.

**1. ÉTICA CONSEQUENCIALISTA E ÉTICA DEONTOLÓGICA**

É moralmente permitido matar uma pessoa para salvar outras cinco? Dilemas morais como esse causam conflitos entre duas principais tradições da ética normativa. O consequencialismo julga a aceitabilidade das ações com base em seus resultados e, portanto, apoia matar um para salvar cinco; *ceteris paribus*, cinco vidas são melhores do que uma. Em contrapartida, a deontologia julga a aceitabilidade das ações de acordo com um conjunto de regras; certas ações (por exemplo, matar) são absolutamente erradas, independentemente de suas consequências.

\* Artigo convidado.

\*\* Assistant Professor of Psychology at Yale University and a Distinguished Research Fellow at the Oxford Centre for Neuroethics. Prior to joining Yale, Dr Crockett was a faculty member at the University of Oxford's Department of Experimental Psychology and a Fellow of Jesus College. She holds a BSc in Neuroscience from UCLA and a PhD in Experimental Psychology from the University of Cambridge and completed a Wellcome Trust Postdoctoral Fellowship with economists and neuroscientists at the University of Zürich and University College London.

\*\*\* Artigo publicado: CROCKETT, Molly, Models of Morality, IN: Trends in Cognitive Sciences, Vol. 17, No. 8, p. 363-366, Julho 2013. Tradução realizada pelo Grupo de Estudos em Direito, Psicologia e Neurociência da Faculdade de Direito de Ribeirão Preto da USP, liderado pelo Prof. Dr. Sérgio Nojiri.

Recentes pesquisas têm mostrado que manipulações experimentais podem influenciar os julgamentos das pessoas em direção ao consequencialismo ou a deontologia, sugerindo que essas perspectivas possuem bases neurais distintas<sup>1</sup>. Um relato influente dessas descobertas postula que os julgamentos deontológicos resultam de processos emocionais automáticos, enquanto que os julgamentos consequenciais resultam de processos cognitivos controlados<sup>2</sup>. Outros argumentam que esta abordagem de processo dual é computacionalmente insuficiente e não pode explicar como cenários hipotéticos são transformados em representações mentais de ações e resultados<sup>3</sup>. A gramática moral universal oferece uma teoria computacional de transformação de problemas, mas carece de uma descrição mecanicista plausível, neurobiologicamente, de como os valores são atribuídos às representações mentais de ações e resultados, e como esses valores são integrados de forma a produzir um julgamento final, consequencialista ou deontológico.

## 2. AVALIAÇÃO BASEADA EM MODELO (*MODEL-BASED*) E SEM MODELO (*MODEL-FREE*)

Os avanços recentes na neurociência oferecem uma nova perspectiva. Avaliações de ações e resultados são orientadas por sistemas de decisão distintos que são psicológica e neurologicamente dissociáveis<sup>4</sup>. O sistema baseado em modelo gera uma árvore de decisões voltadas para o futuro que representa as contingências entre ações e resultados e os valores desses resultados. Ele avalia as ações pesquisando a árvore e determinando quais sequências de ação provavelmente produzirão os melhores resultados. De qualquer maneira, a pesquisa de árvores baseada em modelo é computacionalmente cara e pode tornar-se inutilizável quando as árvores de decisão são excessivamente ramificadas.

O sistema computacionalmente simples e sem modelo não depende de um modelo voltado para o futuro. Em vez disso, avalia ações com base em seus valores anteriormente apreendidos em contextos específicos (situações): boas conjugações de situações são aquelas que produziram resultados desejáveis no passado (por exemplo, abrir a porta), enquanto que as conjugações de situações ruins são aquelas que produziram resultados indesejáveis no passado (por exemplo, empurrar uma pessoa). Como o sistema sem modelo não tem acesso às atuais relações entre ação e resultado, ele é retrospectivo ao invés de prospectivo e pode dar recomendações satisfatórias em situações nas quais as ações tradicionalmente boas conduzem a resultados indesejáveis, ou vice-versa<sup>5</sup>.

Um terceiro sistema, o Pavloviano, promove uma abordagem reflexiva automática e respostas aos estímulos apetitivos e aversivos, respectivamente<sup>6</sup>. Vieses pavlovianos podem influenciar comportamentos guiados por avaliações baseadas em modelo e sem modelo: por exemplo, na transferência aversiva Pavloviana para Instrumental (*Pavlovian-to-instrumental*), as previsões aversivas podem reprimir as ações instrumentais<sup>7</sup>. Os vieses pavlovianos podem também influenciar as próprias avaliações baseadas em modelos: procurar

1 GREENE, Joshua D. The cognitive neuroscience of moral judgment. *The cognitive neurosciences*, v. 4, p. 1-48, 2009.

2 GREENE, Joshua D. The cognitive neuroscience of moral judgment. *The cognitive neurosciences*, v. 4, p. 1-48, 2009.

3 MIKHAIL, John. Universal moral grammar: theory, evidence and the future. *Trends in cognitive sciences*, v. 11, n. 4, p. 143-152, 2007.

4 BALLEINE, Bernard W.; O'DOHERTY, John P. Human and rodent homologues in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, v. 35, n. 1, p. 48, 2010; HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011; WUNDERLICH, Klaus; DAYAN, Peter; DOLAN, Raymond J. Mapping value based planning and extensively trained choice in the human brain. *Nature neuroscience*, v. 15, n. 5, p. 786, 2012; DAYAN, P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. 1068–1074.

5 BALLEINE, Bernard W.; O'DOHERTY, John P. Human and rodent homologues in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, v. 35, n. 1, p. 48, 2010; DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. 1068–1074.

6 DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. 1068–1074.

7 HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011.

uma árvore de decisão pode ser caracterizado como um conjunto de ações internas que podem ser suprimidas por previsões aversivas<sup>8</sup>. Isso equivale a uma “poda” da árvore de decisão, em que a busca de árvores baseada em modelo é reduzida quando um resultado aversivo é encontrado<sup>9</sup>.

Há evidências substanciais de que sistemas baseados em modelos, sem modelos e pavlovianos estão levemente situados em circuitos cerebrais parcialmente distintos, embora os resultados comportamentais provavelmente reflitam sua influência combinada<sup>10</sup>, e evidências recentes sugerem que certas regiões integram avaliações baseadas em modelo e sem modelo<sup>11</sup>. Esses sistemas geralmente chegam a conclusões semelhantes sobre a melhor ação a ser tomada, mas às vezes os resultados divergem. Compreender como esses conflitos são resolvidos é um tópico ativo da pesquisa<sup>12</sup>.

### 3. MODELOS DE MORALIDADE

De modo superficial, o consequencialismo e a deontologia parecem se projetar diretamente para sistemas baseados em modelos e sem modelos, respectivamente. As abordagens consequencialistas e baseadas em modelos avaliam as ações com base em seus resultados, enquanto as abordagens deontológicas e sem modelos avaliam as próprias ações. No entanto, uma análise mais profunda revela que os julgamentos deontológicos provavelmente surgem devido a uma sofisticada interação entre sistemas.

Pense no seguinte problema. No clássico dilema do bonde, um bonde está fora de controle, nos trilhos, em direção a cinco trabalhadores, que irão morrer caso você não faça nada. Você e um homem grande estão de pé em uma passarela acima dos trilhos. Em uma variante deste dilema (alçapão), você pode virar uma alavanca que liberará um alçapão, derrubando o homem grande nos trilhos, onde seu corpo irá parar o bonde. É moralmente permissível virar a alavanca, matar o homem, e salvar os cinco trabalhadores? Em outra variante (empurrão), você pode empurrar o homem grande da passarela para os trilhos, onde novamente seu corpo irá parar o bonde. É moralmente permissível empurrar o homem, matá-lo, e salvar os cinco trabalhadores? Curiosamente, quando as pessoas comuns são confrontadas com esses dilemas, elas são muito menos propensas a aceitar prejudicar uma pessoa para salvar cinco nos casos em que o dano envolve o contato físico com a vítima (como empurrar) do que nos casos em que o dano não envolve contato físico (como liberar um alçapão)<sup>13</sup>, mesmo que esses casos tenham resultados idênticos.

Compreender como diferentes sistemas de decisão avaliam ações e consequências pode trazer uma luz ao problema da divergência entre o empurrão e o alçapão (Box 1). Em consonância com as explicações da gramática moral universal<sup>14</sup>, proponho que o sistema baseado em modelo transforma cenários hipotéticos em uma descrição estrutural de ações e consequências (por exemplo, uma árvore de decisões). Ao examinar a árvore, o sistema baseado em modelo avalia todas as consequências possíveis e recomenda a ação que leva ao melhor resultado.

8 DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. 1068–1074.

9 HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011.

10 BALLEINE, Bernard W.; O'DOHERTY, John P. Human and rodent homologues in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, v. 35, n. 1, p. 48, 2010; HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011; WUNDERLICH, Klaus; DAYAN, Peter; DOLAN, Raymond J. Mapping value based planning and extensively trained choice in the human brain. *Nature neuroscience*, v. 15, n. 5, p. 786, 2012.

11 DAW, Nathaniel D. et al. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, v. 69, n. 6, p. 1204–1215, 2011.

12 DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. 1068–1074.

13 GREENE, Joshua D. The cognitive neuroscience of moral judgment. *The cognitive neurosciences*, v. 4, p. 1-48, 2009; MIKHAIL, John. Universal moral grammar: theory, evidence and the future. *Trends in cognitive sciences*, v. 11, n. 4, p. 143-152, 2007.

14 MIKHAIL, John. Universal moral grammar: theory, evidence and the future. *Trends in cognitive sciences*, v. 11, n. 4, p. 143-152, 2007.

Simultaneamente, o sistema sem modelo avalia ações contextualizadas, atribuindo valores negativos a pares de ação estatal com histórico de reforço negativo (por exemplo, empurrar pessoas)<sup>15</sup>. Uma questão importante é como o sistema sem modelo pode avaliar ações que nunca foram realizadas diretamente (por exemplo, atos violentos). Uma possibilidade é que os valores de ação sejam compreendidos através da observação: um estudo recente mostrou que o aprendizado por observação sem modelo envolve estruturas neurais semelhantes ao do aprendizado experimental sem modelo<sup>16</sup>. Alternativamente, o sistema baseado em modelo poderia treinar o sistema sem modelo através de simulações *off-line*<sup>17</sup>.

Finalmente, o sistema pavloviano pode responder às previsões de sistemas baseados em modelo de resultados aversivos (derivados do enredo textual e representados na árvore de decisão) ou, alternativamente, aos valores aversivos sem modelo atribuídos às ações descritas. Cada sistema “vota” em sua ação preferida, e as escolhas são um produto de sua influência combinada. Podemos explicar a diferença de julgamentos para os casos de empurrão e alçapão considerando que empurrar uma pessoa e virar uma alavanca diferem em termos de suas histórias de reforçamento (*reinforcement histories*) e de resultados próximos esperados, o que, por sua vez, influencia os “votos” dos sistemas sem modelo e pavlovianos (Box 1).

Considere uma segunda característica do julgamento moral: as pessoas prontamente distinguem entre danos causados para a obtenção do fim desejado, e um dano que ocorre como um efeito colateral previsível (Box 2). Essa distinção pode ser vista contrastando o caso do alçapão (descrito acima) com o seguinte caso do trilho alternativo: novamente, um bonde está vindo descontrolado pelos trilhos em direção a cinco trabalhadores, que morrerão caso você não faça nada. Você pode virar uma alavanca que desviará o bonde para uma via de trilhos alternativa, onde um homem grande está parado. É moralmente permissível virar a alavanca, matando o homem grande para salvar os cinco trabalhadores? Apesar de os resultados serem os mesmos, as pessoas consideram que virar a alavanca no caso do alçapão é pior do que no caso do trilho alternativo. Por quê?

No caso do alçapão, o homem grande é usado como um meio: seu corpo serve de instrumento para parar o bonde, evitando que ele atinja os cinco trabalhadores. No caso dos trilhos alternativos, a morte do homem grande é um efeito colateral previsível da ação realizada para salvar os cinco trabalhadores. A estrutura da árvore de decisões é crítica aqui; a distinção entre meios e efeitos colaterais previsíveis pode surgir da poda pavloviana da árvore de decisões (Box 2)<sup>18</sup>.

Visto que as neurobiologias do sistema baseado em modelo, sem modelo e pavloviano são razoavelmente bem delineadas, o panorama atual oferece uma parca explicação para descobertas anteriores. Por exemplo, o córtex pré-frontal medial (mPFC) parece exercer um papel na integração de avaliações do sistema de decisões baseadas em modelo e sem modelo<sup>19</sup>. A proposta de que avaliações de sistema sem modelo contribuem para julgamentos deontológicos, junto com a possibilidade de que o mPFC incorpore valores não baseados em modelo em julgamentos morais, pode explicar duas robustas descobertas na ciência: que casos de contato físico, como o do empurrão, normalmente associados a julgamentos deontológicos, ativam o mPFC<sup>20</sup>; e que pacientes com lesões no mPFC mostram uma tendência menor para julgamentos deontológicos nos

15 CUSHMAN, Fiery et al. Simulating murder: the aversion to harmful action. *Emotion*, v. 12, n. 1, p. 2, 2012; BLAIR, Robert James Richard. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, v. 57, n. 1, p. 1-29, 1995

16 LILJEHOLM, Mimi; MOLLOY, Ciara J.; O'DOHERTY, John P. Dissociable brain systems mediate vicarious learning of stimulus-response and action-outcome contingencies. *Journal of Neuroscience*, v. 32, n. 29, p. 9878-9886, 2012.

17 DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. p. 1068-1074.

18 HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011.

19 BALLEINE, Bernard W.; O'DOHERTY, John P. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, v. 35, n. 1, p. 48, 2010. WUNDERLICH, Klaus; DAYAN, Peter; DOLAN, Raymond J. Mapping value based planning and extensively trained choice in the human brain. *Nature neuroscience*, v. 15, n. 5, p. 786, 2012.

20 GREENE, Joshua D. The cognitive neuroscience of moral judgment. *The cognitive neurosciences*, v. 4, p. 1-48, 2009.

mesmos casos<sup>21</sup>. Previsões pavlovianas aversivas já foram relacionadas ao funcionamento da serotonina<sup>22</sup>; se tais previsões exercem um papel fundamental nos julgamentos deontológicos, aprimorar o funcionamento da serotonina pode melhorar os julgamentos deontológicos, o que já foi, de fato, demonstrado. Finalmente, há evidência de que o stress pode mudar o controle do sistema baseado em modelo para o sistema sem modelo<sup>23</sup>. Minha investigação sugere que o stress deve, de modo similar, promover o julgamento deontológico, como recentemente divulgado<sup>24</sup>.

#### 4. CONSIDERAÇÕES FINAIS

Um benefício dos múltiplos sistemas de decisão é que cada um deles apresenta vantagens em determinadas situações. O sistema baseado em modelo é ótimo para decisões simples. No entanto, quando a árvore de decisões é muito extensa e sua busca se torna computacionalmente inadministrável, mecanismos pavlovianos ou não-baseados em modelo fornecem heurísticas úteis. Talvez os múltiplos sistemas éticos existam pelos mesmos motivos. O consequencialismo fornece um panorama flexível para maximizar o bem e minimizar o mal, mas em face dos resultados incertos ou ambíguos que são comumente encontrados em dilemas do mundo real, as regras deontológicas nos ajudam a evitar problemas. A questão que permanece é a de saber se modelos normativos de tomada de decisão podem moldar éticas normativas.

---

##### Box 1: Explicando a aversão a danos físicos

Um estudo anterior atribuiu os julgamentos deontológicos a processos emocionais automáticos<sup>25</sup>. Faço, aqui, a distinção entre mecanismos racionais sem modelo (mas retreináveis) e mecanismos pavlovianos fixos e ecologicamente racionais, os quais influenciam julgamentos relativos à deontologia em casos em que o dano envolve contato físico.

O sistema não-baseado em modelo avalia ações contextualizadas com base em sua história de reforçamento. As crianças pequenas aprendem através da experiência que as ações que machucam outros fisicamente (por exemplo, bater, empurrar) tem como consequência resultados aversivos (por exemplo, punições, sinais de ansiedade<sup>26</sup>). Simultaneamente, os pais e a sociedade ensinam verbalmente as crianças que é proibido causar danos físicos, alertando sobre as consequências das transgressões. A experiência e a instrução, juntas, permitem que o sistema sem modelo associe valores negativos às ações físicas prejudiciais às pessoas, como em uma classe de algoritmos (chamada Dyna), que complementam o aprendizado experimental de tentativa e erro com o aprendizado hipotético de tentativa e erro<sup>27</sup>. Importante ressaltar que o último método, pelo qual o sistema sem modelo pode ser retreinado por simulações baseadas em modelo, fornece uma via pela qual julgamentos tradicionalmente deontológicos podem ser adaptados a mudanças no ambiente que são detectadas por mecanismos baseados em modelo.

---

21 KOENIGS, Michael et al. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, v. 446, n. 7138, p. 908, 2007.

22 HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011.

HUYS, Quentin J. M. et al. Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, v. 8, n. 3, p. e1002410, 2012.

23 SCHWABE, Lars; WOLF, Oliver T. Stress and multiple memory systems: from 'thinking' to 'doing'. *Trends in cognitive sciences*, v. 17, n. 2, p. 60-68, 2013.

24 YOUSSEF, Farid F. et al. Stress alters personal moral decision making. *Psychoneuroendocrinology*, v. 37, n. 4, p. 491-498, 2012.

25 GREENE, Joshua D. The cognitive neuroscience of moral judgment. *The cognitive neurosciences*, v. 4, p. 1-48, 2009.

26 CUSHMAN, Fiery et al. Simulating murder: the aversion to harmful action. *Emotion*, v. 12, n. 1, p. 2, 2012; BLAIR, Robert James Richard. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, v. 57, n. 1, p. 1-29, 1995.

27 DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. 1068-1074.

Por outro lado, o sistema pavloviano libera respostas para previsões da valência dos estímulos (*valenced stimuli*)<sup>28</sup>; enquanto os valores do estímulo podem ser aprendidos, tendências pavlovianas para aproximar (afastar) estímulos apetitivos (aversivos) são fixos, como reflexos. Para as previsões aversivas, um tipo de resposta pavloviana é a supressão comportamental; essa resposta é ecologicamente racional no sentido de que deixar de agir é geralmente uma boa estratégia quando algumas ações podem produzir resultados aversivos<sup>29</sup>. Previsões aversivas inseridas em dilemas morais poderiam ativar processos pavlovianos que não propiciam respostas ativas, levando a julgamentos caracteristicamente deontológicos. Ações danosas que envolvem contato físico podem gerar previsões aversivas particularmente fortes (por exemplo, expressões de medo, gritos, lesão).

As abordagens computacionais para a tomada de decisões respondem por escolhas, somando valores de ação baseados em modelo, sem modelo e pavloviano e, em seguida, convertem esses valores em probabilidades de ação usando uma função *softmax*<sup>30,31</sup>, tratando essencialmente os três sistemas como especialistas independentes, cada um dos quais “vota” na ação preferida. As diferenças nas histórias de reforçamento de ação estatal (que influenciam os valores livres de modelo) e as previsões aversivas (que influenciam os valores pavlovianos) podem resultar em mais “votos” para a inatividade no cenário de impulso (figura ID) do que no cenário da trapaça (Figura IC), levando a uma maior proporção de julgamentos deontológicos no primeiro do que no último.

28 A valência (valence) é um termo da psicologia muito utilizado na discussão das emoções. Um evento, objeto ou situação pode ter uma atração intrínseca atrativa, “boa” (valência positiva) ou aversiva, “má” (valência negativa).

29 HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011.

30 A função Softmax calcula a distribuição de probabilidades de um evento em ‘n’ diferentes eventos. De modo geral, esta função calculará as probabilidades de cada classe alvo em todas as classes alvo possíveis. Depois, as probabilidades calculadas serão úteis para determinar a classe alvo para as entradas fornecidas.

31 HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011; DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. p. 1068–1074; HUYS, Quentin J. M. et al. Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, v. 8, n. 3, p. e1002410, 2012.

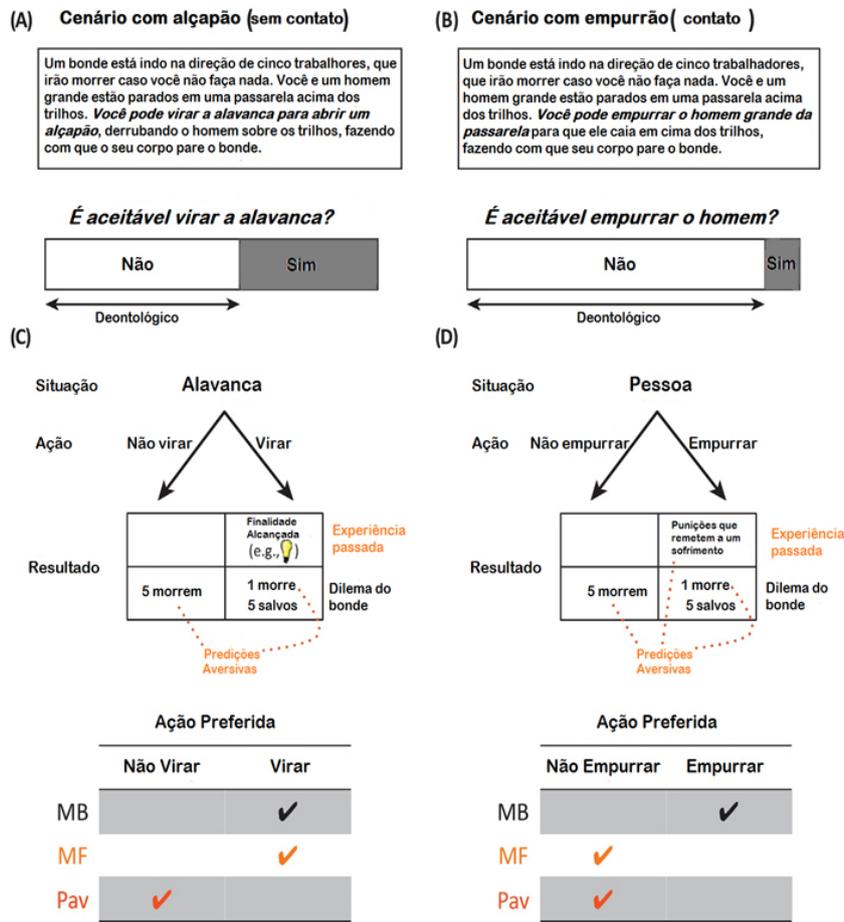


Figura I. (A, B) Exemplo de cenários do bonde e padrões de julgamento típicos (dados adaptados de John Mikhail<sup>32</sup>). (C, D) Diagramas que mapeiam links entre estados, ações e resultados, e tabelas que descrevem as ações preferenciais dos sistemas baseados em modelo (MB), sem modelo (MF) e Pavloviano (Pav).

## Box 2. Explicando a diferença entre meios/efeitos colaterais

Construir e pesquisar por meio de uma árvore de decisão pode ser entendido como um conjunto de ações internas que podem ser suscetíveis aos vieses pavlovianos<sup>33</sup>. Um exemplo é a combinação que leva a estados aversivos, ou uma poda da árvore de decisão<sup>34</sup>. Essa poda é pavloviana, na medida em que é evocada reflexivamente por estados aversivos e persiste mesmo quando contraproducente, impedindo a busca de recompensas que se escondem por detrás de estados aversivos<sup>35</sup>.

Uma séria diferença entre os casos de meios e efeitos colaterais é a posição de dano na árvore de decisão, que tem consequências para a avaliação da ação diante da poda. Considere que a avaliação da ação baseada em modelo integra os resultados de todos os ramos da árvore de decisão, e que a poda pavloviana de ramos que contém consequências aversivas resulta em uma ponderação reduzida de todos os resultados de um ramo podado. Em casos de efeitos colaterais, a poda resulta em um valor maior de ‘virar o interruptor’ reduzindo o

32 MIKHAIL, John. Universal moral grammar: Theory, evidence and the future. *Trends in cognitive sciences*, v. 11, n. 4, p. 143-152, 2007.

33 DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. 1068-1074.

34 HUYS, Quentin J. M. et al. Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, v. 8, n. 3, p. e1002410, 2012.

35 HUYS, Quentin J. M. et al. Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, v. 8, n. 3, p. e1002410, 2012.

peso do resultado aversivo de matar alguém (compare os painéis esquerdo e direito na figura IC). Para salvar os cinco indivíduos, o efeito colateral da morte é incidental, de modo que pode ser podada com segurança, preservando a contribuição de salvar cinco relativamente ao valor geral da ação. No entanto, este tipo de amnésia seletiva para efeitos colaterais negativos não é possível em casos significativos. Aqui, porque o resultado aversivo de matar um é necessário para obter o resultado positivo de salvar cinco, podar o resultado aversivo necessariamente também esconde o resultado positivo. Assim, nos casos de meios, embora a poda reduz o peso do resultado negativo de matar um (assim como em casos de efeitos colaterais), ele também reduz o peso do resultado positivo de salvar cinco (compare os painéis esquerdo e direito na Figura ID). A divergência de julgamentos entre os meios e os casos de efeitos colaterais pode ser explicada pela possibilidade de que, em face da poda, o resultado positivo da economia de cinco contribua menos fortemente para o valor geral da ação nos casos (em que é eliminada) em comparação para evitar casos (em que não é).

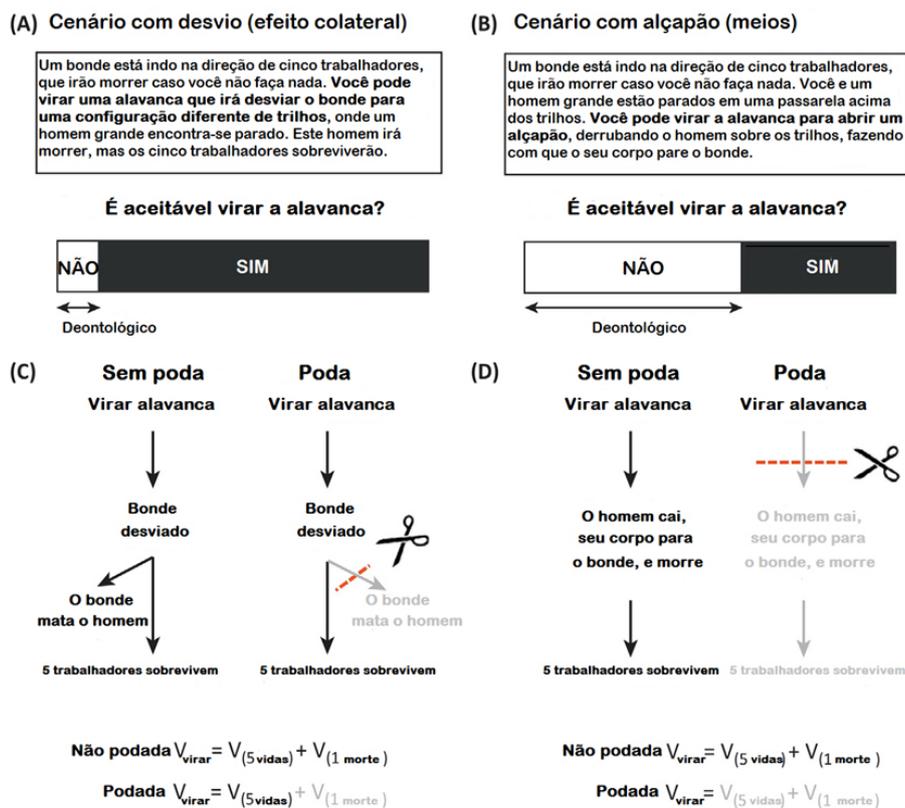


Figura I. (A, B) Exemplo de cenários do bonde e padrões de julgamento típicos (dados adaptados de John Mikhail<sup>36</sup>). (C, D) Árvores de decisão que representam ações e resultados. A poda ocorre no primeiro desfecho aversivo encontrado. O valor geral para virar a alavanca ( $V_{\text{flip}}$ ) é calculado adicionando valores de todos os ramos da árvore: o valor positivo de salvar cinco indivíduos ( $V_{(5 \text{ vidas})}$ ) e o valor negativo de matar um indivíduo ( $V_{(1 \text{ morte})}$ ).

## RECONHECIMENTOS

Este trabalho foi apoiado pelo Wellcome Trust. Agradeço a Peter Dayan, Ray Dolan, Quentin Huys, Michael Moutoussis, Francesco Rigoli, Justin Chumbley e Jenifer Siegel pelas perspicazes discussões e comentários.

36 MIKHAIL, John. Universal moral grammar: Theory, evidence and the future. Trends in cognitive sciences, v. 11, n. 4, p. 143-152, 2007.

## REFERÊNCIAS

- GREENE, Joshua D. The cognitive neuroscience of moral judgment. *The cognitive neurosciences*, v. 4, p. 1-48, 2009.
- MIKHAIL, John. Universal moral grammar: theory, evidence and the future. *Trends in cognitive sciences*, v. 11, n. 4, p. 143-152, 2007.
- BALLEINE, Bernard W.; O'DOHERTY, John P. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, v. 35, n. 1, p. 48, 2010.
- HUYS, Quentin J. M. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS computational biology*, v. 7, n. 4, p. e1002028, 2011.
- WUNDERLICH, Klaus; DAYAN, Peter; DOLAN, Raymond J. Mapping value based planning and extensively trained choice in the human brain. *Nature neuroscience*, v. 15, n. 5, p. 786, 2012.
- DAYAN P. How to set the switches on this thing. *Curr. Opin. Neurobiol.* v. 22, 2012, p. 1068–1074.
- HUYS, Quentin J. M. et al. Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS computational biology*, v. 8, n. 3, p. e1002410, 2012.
- DAW, Nathaniel D. et al. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, v. 69, n. 6, p. 1204-1215, 2011.
- CUSHMAN, Fiery et al. Simulating murder: the aversion to harmful action. *Emotion*, v. 12, n. 1, p. 2, 2012.
- BLAIR, Robert James Richard. A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, v. 57, n. 1, p. 1-29, 1995.
- LILJEHOLM, Mimi; MOLLOY, Ciara J.; O'DOHERTY, John P. Dissociable brain systems mediate vicarious learning of stimulus–response and action–outcome contingencies. *Journal of Neuroscience*, v. 32, n. 29, p. 9878-9886, 2012.
- KOENIGS, Michael et al. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, v. 446, n. 7138, p. 908, 2007.
- CROCKETT, Molly J. et al. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, v. 107, n. 40, p. 17433-17438, 2010.
- SCHWABE, Lars; WOLF, Oliver T. Stress and multiple memory systems: from 'thinking' to 'doing'. *Trends in cognitive sciences*, v. 17, n. 2, p. 60-68, 2013.
- YOUSSEF, Farid F. et al. Stress alters personal moral decision making. *Psychoneuroendocrinology*, v. 37, n. 4, p. 491-498, 2012.

Para publicar na revista Brasileira de Políticas Públicas, acesse o endereço eletrônico [www.rbpp.uniceub.br](http://www.rbpp.uniceub.br)  
Observe as normas de publicação, para facilitar e agilizar o trabalho de edição.