



Civil Procedure Review
AB OMNIBUS PRO OMNIBUS

4

Vieses implícitos e técnicas de automação
decisória: riscos e benefícios

Implicit bias and automated decision-making: risks and benefits

João Paulo Lordelo Guimarães Tavares

PhD. in Law (Federal University of Bahia, Brazil).
Federal Prosecutor in Brazil.

Resumo: O presente artigo tem por objetivo analisar os riscos e benefícios do uso de ferramentas de automação decisória, desenvolvidas por meio de algoritmos, em relação aos vieses implícitos derivados de processos heurísticos. Para tanto, serão inicialmente apresentados conceitos fundamentais à compreensão do tema, a exemplo da classificação dos processos heurísticos extraída da economia comportamental. Em seguida, são expostos os resultados de estudos empíricos que revelam a potencialidade dos vieses cognitivos no campo do processo decisório judicial. Ao final, são apresentados os riscos decorrentes do uso de ferramentas de inteligência artificial – cujos algoritmos podem assimilar os vieses humanos – bem como os possíveis benefícios decorrentes do fato de tais ferramentas executarem tarefas de maneira objetiva, afastadas de processos emocionais.

Palavras-chave: Vieses implícitos. Processos heurísticos. Automação decisória. Algoritmos. Inteligência artificial.

Abstract: This article aims to analyze the risks and benefits of automated decision-making in relation to the implicit biases derived from heuristic processes. For this purpose, concepts that are fundamental to the understanding of the theme will be developed, such as the classification of heuristic processes in behavioral economics. Then, the results of empirical studies that reveal the application of cognitive biases in the field of judicial decision-making are exposed. At the end, the risks arising from the use of artificial intelligence tools - whose algorithms can assimilate human bias - are presented, as well as the possible benefits resulting from the fact that these tools perform tasks in an objective manner, away from emotional processes.

Keywords: Implicit bias. Heuristic processes. Automated decision-making. Algorithms. Artificial intelligence.

Sumário: 1. Introdução: a “descoberta” dos processos mentais heurísticos – 2. Heurísticas e vieses cognitivos na atividade jurisdicional – 2.1 A imparcialidade “consciente” – 2.2 A identificação de vieses raciais na atividade jurisdicional: conhecendo o *Implicit Association Test* (IAT) e outros experimentos – 2.3 A classificação dos vieses cognitivos e sua identificação na atividade jurisdicional – 2.3.1 A sistematização inaugural proposta por Tversky e Kahneman – 2.3.2 Vieses derivados da heurística de representatividade – 2.3.3 Vieses derivados da heurística de disponibilidade – 2.3.4 Vieses derivados da heurística de ancoragem e ajustamento – 2.3.5 Vieses cognitivos particularmente presentes na atividade processual: confirmação, retrospectiva, falácia da conjunção, impossibilidade de esquecimento de provas ilícitas e vieses de decisões sequenciais – 3. O uso da inteligência artificial pode contribuir para reforçar os vieses cognitivos? – 4. O uso da inteligência artificial pode contribuir para afastar os vieses cognitivos? – 5. Conclusão.

1. INTRODUÇÃO: A “DESCOBERTA” DOS PROCESSOS HEURÍSTICOS¹

Ao longo dos debates que ocorreram no processo da eleição presidencial norte-americana de 2016, um importante tema foi lançado aos candidatos: os vieses cognitivos implícitos (“implicit bias”) em abordagens policiais. Questionada a respeito da atuação policial em desfavor da população negra, a então candidata Hillary Clinton, afirmou: “eu acredito que a atuação enviesada é um problema para todos, não apenas para a polícia”.

1. Este artigo é resultado do grupo de pesquisa “Transformações nas teorias sobre o processo e o Direito processual”, vinculado à Universidade Federal da Bahia e cadastrado no Diretório Nacional de Grupos de Pesquisa do CNPQ (dgp.cnpq.br/dgp/espelhogrupo/7958378616800053). Esse grupo é membro fundador da “ProcNet – Rede Internacional de Pesquisa sobre Justiça Civil e Processo contemporâneo” (<http://laprocon.ufes.br/rede-de-pesquisa>).

Dias após, o então candidato Donald Trump acusou a candidata Clinton de ter uma “baixa opinião em relação aos nossos cidadãos”, ao “acusar todo o país, incluindo todas as autoridades de perseguição criminal, de atuar implicitamente de forma enviesada, sugerindo que todos, inclusive nossa polícia, são basicamente racistas e preconceituosos”².

O tema central, no debate em questão, era a chamada política do “stop and frisk”, abordagem policial realizada nas ruas, acusada de focar em jovens negros e hispânicos, alimentando o “feedback loop” nas estatísticas criminais³. Embora não diga respeito, diretamente, à atuação de membros do Poder Judiciário – muito menos às ferramentas de automação decisória –, o assunto parte de premissas que são relevantes também em tal ambiente: os vieses cognitivos.

Há muito, o dogma da ação racional, elemento central da economia clássica, tem sido confrontado por experimentos no campo da psicologia cognitiva⁴ e da economia comportamental (“behavioral economics”)⁵, especialmente a partir da pesquisa seminal realizada na década de 1970, por Amos Tversky e Daniel Kahneman⁶.

2. O debate é registrado por Meagan Biber, destacando o caráter representativo da receptividade do debate sobre preconceitos implícitos na atuação de autoridades públicas. Cf. BOWER, Meagan. Implicit Bias in the Judiciary: Confronting the Problem through Normalization, *Ind. J. L. & Soc. Equal.*, v. 7, 2019, p. 264.
3. O’NEIL, Cathy. Weapons of math destruction: how big data increases inequality and threatens democracy. Nova York: Crown, 2016, p. 93-104.
4. Como registram Dierle Nunes, Natanel Lud e Flávio Quinaud Pedron, “A crença em nossa absoluta racionalidade decisória foi extremamente enfraquecida pelos avanços da psicologia comportamental e por estudos empíricos que demonstram o impacto dos atalhos cognitivos (heurísticas) que nos fazem decidir com deturpações e sem levar em consideração toda a informação relevante” (NUNES, Dierle; LUD, Nathanael; PEDRON, Flávio. *Desconfiando da imparcialidade dos sujeitos processuais*. Salvador: Juspodivm, 2018, p. 10). No mesmo sentido: “Se ha dicho, no sin motivo, que los seres humanos no somos tan racionales como parecemos, en el sentido de que debiéramos ser siempre eficientes y consistentes, y sin embargo no lo somos, a diferencia de las máquinas, que ni se fatigan, ni se distraem o se aburren, ni tienen emociones” (FENOLL, Jordi Nieva. *Inteligencia artificial y proceso judicial*. Marcial Pons: Madrid, Barcelona, Buenos Aires, São Paulo, 2018, p. 43). Em outra passagem, Jordi Nieva Fenoll anota: “Tversky y Kahneman le hicieron un inmenso favor a la humanidad bajándola del pedestal y denunciando que nuestras decisiones no son demasiado complejas. Esto rompía una tendencia en la historia del pensamiento, que, quizá partiendo de un cierto condicionamiento religioso, tendía a elevar al ser humano y sus capacidades intelectuales hasta cotas poco menos que paranormales” (FENOLL, Jordi Nieva. Transfondo psicológico de la independencia judicial. In: FENOLL, Jordi Nieva; OTEIZA, Eduardo (dirs.). *La independencia judicial: um constante asedio*. Marcial Pons: Madri, Barcelona, Buenos Aires, São Paulo, 2019, p. 24).
5. Em resenha publicada por ocasião do lançamento da obra “Thinking, Fast and Slow”, de Daniel Kahneman, Freeman Dyson, matemático e físico inglês, registrou: “the experiment convincingly demolished the central dogma of classical economics” (DYSON, Freeman. How to Dispel Your Illusions. *New York Review of Books*, dez/2011. Disponível em: <http://www.nybooks.com/articles/archives/2011/dec/22/how-dispel-your-illusions/>. Acesso em: 23 ago. 2020).
6. TVERSKY, Amos; KAHNEMAN, Daniel. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, v. 2, n. 5, p. 207-232, 1973.

O escopo do trabalho dos aludidos autores consistiu no processo decisório humano em ambientes de incertezas. Isso porque “muitas decisões são baseadas em crenças relativas à probabilidade de eventos incertos, a exemplo do resultado de uma eleição, a culpabilidade de um réu ou o valor futuro do dólar”⁷. Os experimentos conduzidos por Tversky e Kahneman foram capazes de revelar que, durante tais processos decisórios, muitas pessoas se valem de um limitado número de princípios heurísticos automatizados.

Esses princípios heurísticos consistem em atalhos mentais⁸ ou estratégias inconscientes – heranças do processo evolutivo humano –, meios utilizados na redução da complexidade das tarefas de conhecimento de situações e tomadas de decisões. Eles funcionam como algoritmos cognitivos que permitem que as decisões humanas sejam mais eficientes. Para tanto, são realizados processos inconscientes de categorização de objetos, pessoas e suas ocorrências em grupos ou tipos.

Além de adaptativas, as heurísticas são inegavelmente úteis, contribuindo para uma maior eficiência nas decisões diárias, notadamente aquelas que integram o chamado “Sistema 1” de pensamento, em que os julgamentos e escolhas são mais intuitivos, experimentais e automatizados⁹.

Apesar disso, elas podem resultar em erros severos e sistemáticos¹⁰, conduzindo a falácias, vieses e ilusões. Cuida-se do “preço que pagamos para tamanha eficiência”¹¹.

Some-se a isso o fato de que, cada vez mais, no campo do Direito, decisões humanas são delegadas a ferramentas de inteligência artificial desenvolvidas a partir do uso de algoritmos. No Brasil, além do projeto VICTOR – do Supremo Tribunal Federal

7. TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. *Science*, v. 185, p. 1.124-1.131, 1974.
8. Nesse sentido: “Heuristics are cognitive shortcuts, or rules of the thumb, by which people generate judgments and make decisions without having to consider all the relevant information, relying instead on a limited set of cues that aid their decision making” (PEER, Eyal; GAMLIEL, Eyal. Heuristics and Biases in Judicial Decisions. *Court Review*, v. 49, n. 2, p. 114-119, 2013). Cf. JONES, Craig. The Troubling New Science of Legal Persuasion: Heuristics and Biases in Judicial Decision-Making. *Advocates Quarterly*, v. 41, p. 49-122, 2013; JONES, Craig; RANKIN, Micah B. E. Justice as a Rounding Error? Evidence of Subconscious Bias in Second-Degree Murder Sentences in Canada. *Osgoode Digital Commons*, v. 10, n. 81, 2014.
9. KAHNEMAN, Daniel. *Rápido e Devagar*. Tradução: Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012.
10. Cf. SUNSTEIN, Cass R. *Behavioral Law and Economics*. Cambridge: Cambridge University Press, 2000; NUNES, Dierle; LUD, Nathanael; PEDRON, Flávio. *Desconfiando da imparcialidade dos sujeitos processuais*. Salvador: Juspodivm, 2018; FENOLL, Jordi Nieva. Transfondo psicológico de la independencia judicial. In: FENOLL, Jordi Nieva; OTEIZA, Eduardo (dirs.). *La independencia judicial: um constante asedio*. Marcial Pons: Madri, Barcelona, Buenos Aires, São Paulo, 2019; COSTA, Eduardo José da Fonseca. *Levando a imparcialidade a sério: proposta de um modelo interseccional entre direito processual, economia e psicologia*. Salvador: Juspodivm, 2018.
11. NEGOWETTI, Nicole. E. Judicial Decisionmaking, Empathy and the Limits of Perception. *Akron Law Review*, v. 47, 2014.

–, podem ser citados os robôs Alice, Sofia e Monica, que têm sido empregados pelo Tribunal de Contas da União para, respectivamente: a) examinar milhares de editais de licitação e atas de preços em busca de fraudes e irregularidades; b) analisar e sugerir aprimoramentos em relatórios internos; e c) acompanhar todas as compras públicas, inclusive as decorrentes de contratação direta. Em razão do sucesso de tais instrumentos, seu uso foi transportado para outros órgãos de controle, a exemplo dos Tribunais de Contas estaduais, Ministério Público, Polícia Federal e Controladoria Geral da União¹².

Daí decorrem duas preocupações fundamentais, de ordem democrática: a) a necessidade de se assegurar o direito fundamental de acesso à informação a todos os cidadãos sujeitos a tais mecanismos; e b) o dever do Estado de prevenir e corrigir as distorções decorrentes do desenvolvimento, alimentação ou uso inadequado das ferramentas, evitando que as decisões produzidas sejam enviesadas e discriminatórias.

2. HEURÍSTICAS E VIESES COGNITIVOS NA ATIVIDADE JURISDICIONAL

2.1 A imparcialidade “consciente”

Uma famosa anedota americana retrata três árbitros de beisebol questionados sobre como eles conduzem um jogo. Um diz: “eu faço de acordo com o que vejo”. Outro afirma: “eu apito e conduzo conforme as coisas são”. Um último responde: “não existe ponto ou falta, até eu apitar”¹³.

Nesse cenário anedótico, enquanto o primeiro juiz reconhece sua condição de observador humano – e, portanto, imperfeito –, os demais se veem como infalíveis, capazes de conduzir com base em seus méritos objetivos.

De forma semelhante, o *Justice* Cardozo – que atuou na Suprema Corte dos Estados Unidos entre 1932 e 1938 – já se manifestou no sentido de que os juízes teriam a capacidade de “se afastar de qualquer influência pessoal ou que advenha de uma situação particular que lhe é apresentado, fundamentando suas decisões em elementos de natureza objetiva”¹⁴.

12. FERRARI, Isabela; BECKER, Daniel; WOLKART, Erik Navarro. Arbitrium ex machina: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos. *Revista dos Tribunais*, v. 995, p. 635-655, set./2018.

13. A anedota narrada consiste numa tradução livre e adaptada da introdução de artigo de autoria de Eyal Peer e Eyal Gamliel (PEER, Eyal; GAMLIEL, Eyal. Heuristics and Biases in Judicial Decisions. *Court Review*, v. 49, n. 2, p. 114-119, 2013).

14. OAKES, Anne Richardson; DAVIES, Haydn. Process, Outcomes and the Invention of Tradition: The Growing Importance of the Appearance of Judicial Neutrality. *Santa Clara Law Review*, v. 51, n. 2, 2011.

Essa clássica noção de imparcialidade deriva da crença de que os julgadores são capazes de, conscientemente, afastar-se de vieses pessoais. Muitos juízes acreditam nessa possibilidade, tal como exigido em códigos de conduta¹⁵.

Pesquisas recentes demonstram o contrário. Muitos juízes ainda desconhecem os vieses cognitivos implícitos e, em muitos casos – a exemplo das questões relativas a preconceitos raciais –, tratam o assunto como tabu. No atual cenário das questões éticas relativas à prestação jurisdicional, julgadores continuam livres para atuar de forma despreocupada quanto aos vieses implícitos que habitam o inconsciente.

Esses vieses podem trazer consequências diversas, que variam desde a má apreciação das provas de um caso até associações estereotipadas.

Apesar disso, um tema que raramente é abordado, quando se trata da independência judicial, diz respeito à psicologia dos juízes¹⁶. De fato, poucos profissionais jurídicos – sejam administradores públicos, advogados, juízes ou membros do Ministério Público – atentam ao fato de que o pensamento humano não é habitualmente muito complexo, valendo-se de parâmetros “pseudoestatísticos”¹⁷ para a tomada de decisões.

Ao se deparar com uma tarefa decisória, o ser humano frequentemente tende a se socorrer da própria memória – recordando aquilo que mais frequentemente considera exitoso – ou simplesmente defender uma opinião prévia, o que pode estar fundamentado em um cálculo mental equivocados.

Estudos empíricos demonstram que os juízes costumam ser mecânicos na maioria das ocasiões, classificando os processos de forma sistemática, de modo a facilitar o seu trabalho. Motivações previamente elaboradas costumam ser replicadas em casos similares, com as adaptações necessárias. Esse processo é absolutamente natural e objetiva imprimir eficiência na atividade jurisdicional.

15. Conforme estabelecido no art. 8º do Código de Ética da Magistratura, editado pelo Conselho Nacional de Justiça, “O magistrado imparcial é aquele que busca nas provas a verdade dos fatos, com objetividade e fundamento, mantendo ao longo de todo o processo uma distância equivalente das partes, e evita todo o tipo de comportamento que possa refletir favoritismo, predisposição ou preconceito” (Disponível em: <https://www.cnj.jus.br/codigo-de-etica-da-magistratura/>. Acesso em: 23 ago. 2020). De igual modo, o Canon 3 do Código Modelo de Conduta Judicial (*Model Code of Judicial Conduct*) dos Estados Unidos destaca: “The judge should perform those duties with respect for others, and should not engage in behavior that is harassing, abusive, prejudiced, or biased. The judge should adhere to the following standards” (Disponível em: <https://www.uscourts.gov/judges-judgeships/code-conduct-united-states-judges#c>. Acesso em 23 ago. 2020).
16. São exemplos de autores que abordam a temática em questão: POSNER, Richard. *How judges think*. Harvard Press, 2010; FENOLL, Jordi Nieva. Transfondo psicológico de la independencia judicial. In: FENOLL, Jordi Nieva; OTEIZA, Eduardo (dirs.). *La independencia judicial: um constante asedio*. Marcial Pons: Madri, Barcelona, Buenos Aires, São Paulo, 2019; NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, v. 285, p. 421-447, nov./2018.
17. Cf. KAHNEMAN, Daniel. TVERSKY, Amos. *On the study of statistical intuitions*. *Cognition*, 1982, p. 123.

Ocorre que, não raramente, os órgãos públicos se deparam com um grande volume de casos para serem analisados, o que resulta em um incremento da mecanização – e até mesmo automatização – do processo de classificação e decisão. Salvo nos casos mais complexos, o tempo dedicado a cada processo é bastante limitado, o que amplifica a influência de processos heurísticos de pensamento.

Tais considerações revelam a necessidade de que a ciência jurídica se ocupe também do estudo dos processos heurísticos nas decisões jurídicas, introduzindo, na medida do possível, as possibilidades de desvios (“debiasing”) no campo do devido processo legal.

2.2 A identificação de vieses raciais na atividade jurisdicional: conhecendo o *Implicit Association Test* (IAT) e outros experimentos

Um importante experimento, descrito, entre outros, em artigo de Megan Biwer¹⁸, consiste no estudo desenvolvido por pesquisadores da Universidade de Harvard, por meio do *Implicit Association Test* (IAT). Cuida-se de relevante ferramenta, amplamente reconhecida para a identificação de vieses implícitos.

Em síntese, o IAT tem objetivo medir a força de associações inconscientes por meio da comparação do tempo que um indivíduo gasta para fazê-las. A título exemplificativo, considere-se a necessidade de leitura de duas listas: a primeira, contendo nomes de cores escritos na cor respectiva (ex.: a palavra “azul” escrita na tonalidade azul); a segunda, contendo nomes de cores escritos em tonalidades aleatórias (ex.: a palavra “vermelho” escrita em tom verde). Os experimentos revelam que a leitura da primeira lista ocorre de forma consideravelmente mais curta. A segunda lista, por sua vez, toma mais tempo para ser lida. Essa diferença de tempo consiste no importante elemento variável medido pelo IAT.

O “Race IAT” tem por objetivo identificar vieses implícitos em desfavor de pessoas de variadas etnias. Como relata Rachlinski, inicialmente pesquisadores apresentam aos participantes uma tese de computador com as palavras “Branco ou Bom”, no canto esquerdo superior, e “Negro ou Mau”, no canto direito superior. Em seguida, são apresentados, individualmente, quatro tipos de estímulos no centro da tela: rostos de pessoas negras, rostos de pessoas brancas, palavras positivas e palavras negativas. Os participantes são instruídos a pressionar uma tecla na esquerda do computador quando aparecer uma face branca ou uma palavra positiva, e a pressionar uma tecla na direita quando uma face negra ou uma palavra negativa aparecem. Em seguida, o

18. BIWER, Meagan. *Implicit Bias in the Judiciary: Confronting the Problem through Normalization*, *Ind. J. L. & Soc. Equal*, v. 7, 2019, p. 268. Cf. também: RACHLINSKI, Jeffrey J.; JOHNSON, Sheri; WISTRICH, Andrew J.; GUTHRIE, Chris. *Does unconscious racial bias affect trial judges?* *Cornell Law Faculty Publications*, 2009. Disponível em: <https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1691&context=facpub>. Acesso em: 30 ago. 2020.

computador é programado para alternar a localização das palavras “Bom” e “Mau”, de modo que as palavras “Branco ou Mau” apareçam no canto superior esquerdo da tela, e as palavras “Negro ou Bom” apareçam no canto superior direito. Os participantes também repetem o mesmo processo com novos pares opostos¹⁹.

O que os pesquisadores concluíram, com esse experimento, foi que muitos americanos demonstram uma forte “preferência branca” no desempenho da tarefa, na medida em que o tempo gasto com o desempenho da primeira tarefa revelou-se muito mais rápido, associando-se as palavras “Branco” a “Bom” e “Negro” a “Mau”. Aproximadamente 70% (setenta por cento) das pessoas que fizeram o teste do IAT demonstraram preferências raciais, incluindo participantes negros.

Esses dados são corroborados por outros estudos.

Outros experimentos similares, fora do IAT, revelaram que pessoas brancas tendem a ativar a região do cérebro associada com o medo, ao enxergarem rostos de pessoas negras, o que tem levantado sérias preocupações quanto ao encarceramento de pessoas latinas e negras nos Estados Unidos²⁰.

Pesquisas realizadas nos Estados Unidos informam que brancos associam implicitamente os negros (e não os brancos) com imagens de primatas não humanos e, quanto mais facilmente o fazem, mais inclinados estão a endossar a violência policial contra os negros. Nesse sentido, em estudos de casos ocorridos na Flórida, Geórgia e Pensilvânia, pesquisadores concluíram que presidiários com características faciais mais fenotipicamente negras (lábios mais grossos, narizes mais largos, etc.), em circunstâncias fáticas similares, receberam sentenças mais longas do que aqueles com menos características fenotípicas negras.

Conforme dados apresentados pelo Georgia Department of Corrections²¹, não apenas os negros recebem sentenças 4,25% mais altas do que os brancos, mas os

19. RACHLINSKI, Jeffrey J.; JOHNSON, Sheri; WISTRICH, Andrew J.; GUTHRIE, Chris. Does unconscious racial bias affect trial judges? *Cornell Law Faculty Publications*, 2009. Disponível em: <https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1691&context=facpub>. Acesso em: 30 ago. 2020.

20. Nesse sentido: “Even more, the social scientific literature underscores that blacks (and not whites) are implicitly perceived as a threat and hostile, which is particularly important to consider in the context of racial disparities in death penalty sentencing—the punishment that should be reserved for the most “threatening” people or horrific crimes. For example, on neuroimaging measures, whites show more activation in the region of the brain associated with fear when they view black faces. Where there is a confrontation between people, whites also implicitly construe those interactions as more aggressive and hostile when there is a black perpetrator” (PARKS, Gregory S.; DAVIS, Andre M. Confronting Implicit Bias: An Imperative for Judges in Capital Prosecutions. *Human Rights Magazine*, v. 42, n. 2, 2016. Disponível em: https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/2016-17-vol-42/vol-42-no-2---the-death-penalty--how-far-have-we-come-/confronting-implicit-bias--an-imperative-for-judges-in-capital-p/. Acesso em: 30 ago. 2020).

21. Cf. BURCH, Traci. Skin Color and the Criminal Justice System: Beyond Black-White Disparities in Sentencing. *Journal of Empirical Legal Studies*, v. 12, p. 395-420, 2015.

negros de pele média e escura também recebem sentenças 4,8% mais altas do que as dos brancos. Por outro lado, os negros de pele clara recebem sentenças quase da mesma severidade que os brancos. Pesquisadores também concluíram que, em casos que envolvem uma vítima branca, quanto mais fenotipicamente negro o réu era considerado, maior era a probabilidade de ele ser condenado à morte²².

De igual modo, outros estudos apontam que juízes tendem a fixar fianças consideravelmente mais elevadas, em acréscimo de 25% (vinte e cinco por cento), bem como punições cerca de 12% (doze por cento) mais longas em desfavor de réus negros em situações similares a réus brancos²³.

Existem, portanto, fortes experimentos que demonstram, empiricamente, a existência de vieses implícitos capazes ensejar a tomada de decisões irracionais, muitas delas em desfavor de minorias²⁴.

Recentes estudos demonstram até mesmo fatores aparentemente irrelevantes são considerados em processos heurísticos decisórios. A título de exemplo, juízes levados a refletir sobre sua própria morte apresentaram maior probabilidade de proferir decisões mais conservadoras. Do mesmo modo, juízes em tribunais que se encontravam temporalmente mais distantes da sua última refeição apresentaram

-
22. PARKS, Gregory S.; DAVIS, Andre M. Confronting Implicit Bias: An Imperative for Judges in Capital Prosecutions. *Human Rights Magazine*, v. 42, n. 2, 2016. Disponível em: https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/2016-17-vol-42/vol-42--no--2---the-death-penalty--how-far-have-we-come-/confronting-implicit-bias--an-imperative-for-judges-in-capital-p/. Acesso em: 30 ago. 2020). No mesmo sentido: EBERHARDT, Jennifer L; DAVIES, Paul G.; PURDIE-VAUGHNS, Valerie J.; JOHNSON, Sheri Lynn. Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes. *Psychological Science*, v. 17, p. 383-386, 2006.
23. Cf. WALDFOGEL, Joel; AYERS, Ian. A Market Test for Race Discrimination in Bail Setting. *Stanford Law Review*, v. 46, p. 987-1048, 1994.
24. Fora do sistema de justiça, estudo conduzido por pesquisadores da Universidade Federal da Bahia e da Universidad Complutense de Madrid, envolvendo 101 participantes brasileiros e 138 espanhóis, revelou a influência de categorias naturalistas (como a cor da pele) na construção de estereótipos sociais. Segundo os pesquisadores, o estudo destacou a “categoria raça” como aquela que sofreu o maior grau de essencialização, o que ajuda a compreender porque “a categorização de um grupo em função do fenótipo é um elemento poderoso na racialização das relações intergrupais, em cuja base as desigualdades sociais obtêm legitimidade” (PEREIRA, Marcos Emanuel; ÁLVARO, José Luis; OLIVEIRA, Andréia C. Oliveira; DANTAS, Gilcimar. Estereótipos e essencialização de brancos e negros: um estudo comparativo. *Psicologia & Sociedade*, v. 1, p. 144-153, 2011). Essas conclusões vão ao encontro de recentes estudos que identificaram vieses raciais implícitos na contratação de trabalhadores e no aprisionamento de adolescentes negros por agentes policiais nas ruas (Cf. ROOTH, Dan-Olof. Implicit Discrimination in Hiring: Real World Evidence. Institute for the Study of Labor (IZA), 2007. Disponível em: <http://ftp.iza.org/dp2764.pdf>. Acesso em 30 ago. 2020; GOFF, Phillip Atiba; JACKSON, Matthew Christian; DI LEONE, Brooke Allison Lewis; CULOTTA, Carmen Marie; DITOMASSO, Natalie Ann. The Essence of Innocence: Consequences of Dehumanizing Black Children. *Journal of Personality and Social Psychology*, v. 106, n. 4, p. 526-545, 2014. Disponível em: <https://www.apa.org/pubs/journals/releases/psp-a0035663.pdf>. Acesso em: 30 ago. 2020).

maior probabilidade de confirmar a decisão recorrida. Até mesmo o momento das alegações finais e as vestimentas dos advogados foram identificados como fatores relevantes no processo decisório²⁵.

2.3 A classificação dos vieses cognitivos e sua identificação na atividade jurisdicional

2.3.1 A sistematização inaugural proposta por Tversky e Kahneman

Para além da identificação de vieses relacionados a gênero e raça no exercício da atividade jurisdicional, o processo de tomada de decisões costuma se valer, não raramente, de processos heurísticos que, com o objetivo inconsciente de reduzir a complexidade da tarefa, conduzem a erros lógicos sistemáticos.

Por conseguinte, é possível identificar uma série de vieses (falácias, ilusões, erros lógicos etc.) que acompanham os processos judiciais em suas diversas fases: admissibilidade, produção de provas, sentença, fase recursal etc.

As classificações de vieses derivados de processos heurísticos, dentro e fora do sistema de justiça, apresentam nomenclaturas diversas. Tradicionalmente, contudo, costuma-se partir dos três processos heurísticos probabilísticos propostos por Amos Tversky e Daniel Kahneman: representatividade (“representativeness”), disponibilidade (“availability”) e ancoragem-ajustamento (“adjustment and anchoring”).

2.3.2 Vieses derivados da heurística de representatividade

O processo heurístico de representatividade tem lugar quando se exige de alguém a tarefa de enquadrar um indivíduo, um objeto ou um evento em uma determinada categoria. Em síntese, questiona-se a probabilidade de um objeto ou indivíduo A pertencer à categoria B, ou mesmo a probabilidade de um evento A derivar do processo B²⁶.

Desse processo heurístico, um problema (viés) comum é que “as pessoas tendem a se focar em determinadas características dos elementos, sem perceber o quão comum essas características de fato são, nos outros”²⁷. Assim, algumas propriedades do objeto, pessoa ou evento em análise acabam sendo superdimensionadas, tomados como representativos das categorias. O resultado é a construção de uma espécie de “probabilidade subjetiva”, contrária à real representatividade – e, portanto, probabilidade real – de um determinado evento.

25. JONES, Craig E. The Troubling New Science of Legal Persuasion: Heuristics and Biases in Judicial Decision-Making. *Advocates’ Quarterly*, v. 41, n. 1, 2013.

26. TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. *Science*, v. 185, 1974, p. 1124.

27. VITORELLI, Edilson. *Processo civil estrutural: teoria e prática*. Salvador: Juspodivm, 2020, p. 376.

Essa má apreciação dos dados pode se manifestar de diferentes formas: insensibilidade a outros dados probabilísticos, insensibilidade ao tamanho da amostra (em violação à regra estatística básica segundo a qual as oscilações de resultado tendem a ser maiores em amostras menores), má representação da aleatoriedade (com a tendência de apego a eventos aleatórios), insensibilidade à previsibilidade, equívocos na apreciação de dados em regressão etc.²⁸

A heurística de representatividade, quando relacionada à identificação ou classificação de pessoas, costuma recorrer ao terreno perigoso dos estereótipos. De forma exemplificativa, em um juízo sumário, se um indivíduo busca saber a probabilidade de a pessoa X ser um engenheiro, o caminho natural será a identificação, nessa pessoa X, de atributos representativos do estereótipo de engenheiro. Essa forma de raciocínio costuma produzir muitos equívocos, na medida em que são considerados fatores não causais.

Em um curioso experimento conduzido por Tversky e Kahneman, indivíduos foram apresentados a uma breve descrição da personalidade de pessoas aleatoriamente retiradas de uma amostra de 100 (cem) profissionais, entre engenheiros ou advogados. Os participantes foram então questionados, a partir de cada descrição de cada pessoa, a respeito da probabilidade de se tratar de um engenheiro ou advogado. Em uma condição experimental, os participantes foram previamente informados que, de toda a amostra de 100 (cem) pessoas, 70 (setenta) eram engenheiros e 30 (trinta), advogados. Em outra, o oposto – 70 (setenta) advogados e 30 (trinta) engenheiros. Em uma clara violação do teorema de Bayes, os participantes produziram essencialmente os mesmos juízos de probabilidade. Ignorando por completo os percentuais previamente informados, julgaram as pessoas com base exclusivamente os estereótipos de engenheiro ou advogado²⁹.

Na seara jurídica, não há dúvidas de que os variados profissionais (juízes, advogados, membros do Ministério Público etc.) tendem a classificar os casos a partir da recordação de problemas mais comuns em cada campo. O recurso à representatividade faz trazer à memória precedentes judiciais e dados análogos de casos anteriores em que atuaram. Os problemas se intensificam quando, em lugar de engenheiro ou advogado, o juízo a ser realizado por um jurista diz respeito à probabilidade de uma pessoa X ter elevada periculosidade – qualidade buscada, por exemplo, na aplicação de medidas de segurança ou medidas cautelares como a prisão preventiva. Consciente ou inconscientemente, até mesmo o raciocínio sobre se a pessoa Y cometeu um determinado delito costuma partir de uma “probabilidade subjetiva” que, segundo informam os experimentos citados, acabam por considerar

28. TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. *Science*, v. 185, 1974, p. 1124.

29. TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. *Science*, v. 185, 1974, p. 1124.

características que transcendem a conduta e adentram a *persona* do acusado (etnia, nacionalidade, aparência física etc.).

Os vieses derivados da heurística de representatividade também podem ser bastante prejudiciais no âmbito do processo civil, em especial no que diz respeito à produção de provas. A tendência que as pessoas têm de extrair semelhanças entre indivíduos, condições e exemplos, fazendo inferências de causalidade a partir da similaridade (ou seja, representatividade) do que se lhes apresenta repetidamente, pode conduzir a conclusões equivocadas. Em especial, eventos similares tendem a ser causalmente relacionados, enquanto eventos causais – porém distintos – tendem a ser percebidos como não relacionados³⁰.

2.3.3 Vieses derivados da heurística de disponibilidade

Pela heurística de disponibilidade (“availability”), as pessoas valoram a frequência ou a probabilidade de um evento ter ocorrido em função da facilidade com que recordações são trazidas à mente. Ordinariamente, a heurística de disponibilidade é uma ferramenta bastante útil, na medida que eventos em larga escala tendem a ser rememorados mais facilmente e mais rapidamente, em comparação com eventos de menor probabilidade.

O problema surge quando eventos raros – ou menos comuns – são percebidos como se fossem mais frequentes, pelo simples fato de serem recordados com facilidade. Basta comparar o medo que as pessoas costumam associar a viagens de avião, normalmente mais elevado que o medo de viagens de carro, embora estas sejam indiscutivelmente mais perigosas.

Um pequeno experimento conduzido por Tversky e Kahneman demonstra isso. Participantes ouviram o conteúdo de listas de nomes de personalidades famosas de ambos os sexos e, em seguida, foram questionados sobre se as listas conteriam mais nomes masculinos ou femininos. Diversas listas foram apresentadas a diferentes grupos de pessoas. Em algumas das listas, os homens eram relativamente mais famosos que mulheres; e outras, o inverso. Em cada uma das listas, os participantes julgaram, equivocadamente, que a classe (sexo) que continha personalidades mais famosas era também a mais numerosa³¹.

Outro experimento, conduzido por John Carroll, demonstrou que pessoas que eram solicitadas a imaginar a vitória de um determinado candidato nas eleições tinham a tendência de, em entrevista posterior, considerá-lo com mais chances de vitória na

30. VITORELLI, Edilson. *Processo civil estrutural: teoria e prática*. Salvador: Juspodivm, 2020, p. 376.

31. TVERSY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. *Science*, v. 185, 1974, p. 1124.

disputa. O mero exercício imaginativo anterior, por se tratar de evento recente na memória, foi capaz de influenciar o julgamento probabilístico³².

No campo jurídico, esse viés incrementa o risco de julgadores dimensionarem excessivamente a probabilidade de ocorrência de um determinado evento, com fundamento em uma experiência pessoal isolada, mas traumática. É possível supor, por exemplo, que pessoas que, de alguma forma, foram sujeitas a algum tipo de violência policial, tendem a dar mais credibilidade a esse argumento, ainda que desacompanhado de evidências mínimas.

Um outro exemplo, trazido por Edilson Vitorelli, diz respeito ao fato de juízes serem mais propensos a conceder direitos que sejam fundados em histórias individuais dramáticas, a exemplo de demandas individuais relacionadas a tratamentos médicos, do que pretensões embasadas em dados estatísticos sólidos³³.

No âmbito da tutela cautelar, Jordi Nieva Fenoll registra que este tipo de viés é bastante frequente na percepção do risco (*periculum in mora*). Essa percepção tende a ser mais intensa quando o julgador houver denegado previamente a tutela em caso similar que resultou em prejuízo ao réu, em processo civil, ou mesmo possibilitou a reincidência que se buscava evitar, em processo penal³⁴.

2.3.4 Vieses derivados da heurística de ancoragem e ajustamento

Pela heurística de ancoragem e ajustamento, em muitas situações, as pessoas fazem estimativas a partir de um valor inicial, que é ajustado para produzir a resposta final. O valor inicial – concebido como um ponto de partida – pode ser sugerido pela formulação do problema ou mesmo resultar de um processo de cálculo parcial³⁵.

Os vieses são recorrentes, na medida em que os ajustamentos costumam ser insuficientes, de modo que os pontos de partidas produzem diferentes resultados viciados a partir dos valores iniciais.

Em um conhecido experimento, participantes foram solicitados a estimar determinadas quantidades aleatórias (por exemplo, o número de países africanos que integram a Organização das Nações Unidas). Antes de cada questionamento, um número entre 0 (zero) e 100 (cem) era previamente selecionado de forma aleatória, por meio de uma roleta. Primeiramente, os participantes eram instruídos a informar

32. CARROLL, John S. The effect of imagining an event on expectations for the event: an interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology*, v. 14, n. 1, p. 88-96, 1978.

33. VITORELLI, Edilson. *Processo civil estrutural: teoria e prática*. Salvador: Juspodivm, 2020, p. 375.

34. FENOLL, Jordi Nieva. *Inteligencia artificial y proceso judicial*. Marcial Pons: Madrid, Barcelona, Buenos Aires, São Paulo, 2018, p. 49.

35. TVERSY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. *Science*, v. 185, 1974, p. 1128.

se o valor questionado seria maior ou menor que o valor aleatoriamente indicado na roleta. Em seguida, eram solicitados a estimar o percentual. A diferentes grupos foram fornecidos diferentes números para cada quantidade questionada. Ao final, foi possível concluir que os números arbitrários produziram efeitos sobre as estimativas. Por exemplo, a média estimada para o percentual de países africanos na ONU foi de 25 (vinte e cinco) e 45 (quarenta e cinco) para grupos que retiraram 10 (dez) e 45 (quarenta e cinco), respectivamente, como pontos de partida arbitrários indicados na roleta³⁶.

Esse processo de construção de uma conclusão mediante a assimilação de dados previamente considerados como pontos de partida pode produzir variados efeitos no campo jurídico.

Sentenças criminais condenatórias, por exemplo, possuem expressão numérica quanto à fixação da pena, havendo estudos indicativos da possibilidade de juízes serem influenciados por outros valores previamente apresentados³⁷.

Pesquisas conduzidas por Mussweiler, Englich e Strack revelam que mesmo juízes americanos experientes são sujeitos ao efeito de ancoragem, sendo influenciados por dados anteriormente sugeridos pelo Ministério Público ou pela mídia.

Como revelam os autores, juízes e jurados têm a relevante tarefa de decidir a respeito da condenação de alguém em um ambiente de incerteza e quadro probatório controverso. Ao contrário do que pode parecer em uma primeira análise, a existência de regras e princípios jurídicos é capaz de minimizar as influências de fatores irrelevantes no processo decisório, mas não de eliminá-las. No contexto do processo civil, experimentos sugerem que a quantia requerida pela parte autora a título de danos materiais e punitivos pode servir como valor-âncora. Em outras palavras, quanto maior o valor requerido pela parte autora, maior a probabilidade de uma indenização mais elevada³⁸.

Efeitos similares têm sido observados no campo do processo penal, em que os valores numéricos requeridos ou sugeridos ao Poder Judiciário exercem forte influência sobre as decisões judiciais. Em relevante estudo publicado em 2001, Englich e Mussweiler já haviam concluído que, em casos envolvendo o delito de estupro,

36. TVERSY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. *Science*, v. 185, 1974, p. 1128.

37. Nesse sentido, conferir: PEER, Eyal; GAMLIEL, Eyal. Heuristics and Biases in Judicial Decisions. *Court Review*, v. 49, n. 2, p. 114-119, 2013; ENGLISH, Birt; MUSSWEILER, Thomas. Sentencing Under Uncertainty: Anchoring Effects in the Courtroom. *Journal of Applied Social Psychology*, v. 31, n. 7, p. 1535-1551, 2001; STEIN, Christopher; DOUIN, Michelle. Cognitive Bias in the Courtroom: Combating the Anchoring Effect in Criminal Sentencing. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2991611. Acesso em: 31 ago. 2020.

38. ENGLISH, Birt; MUSSWEILER, Thomas; STRACK, Fritz. Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making. *Personality and Social Psychology Bulletin*, v. 32, p. 188-200, 2006.

juízes são altamente influenciados pelos termos do pedido condenatório formulado pelo Ministério Público³⁹.

Em 2006, agora também com a participação de Strack, referidos autores realizaram experimento mais profundo, dividido em três estudos, nos quais juízes foram previamente expostos a uma potencial conclusão antes de fazer a sua própria decisão. No primeiro estudo, a potencial sentença-modelo derivou de uma fonte que, em termos normativos, deveria ser considerada irrelevante para a decisão judicial: um jornalista. No segundo, os participantes foram informados que o pedido condenatório havia sido formulado pelo Ministério Público de forma aleatória. No terceiro, os próprios participantes foram informados dos termos do pedido condenatório aleatoriamente, lançando um conjunto de dados.

Em todos os três estudos, os participantes expostos a valores de ancoragem mais elevados produziram decisões consideravelmente mais elevadas, a revelar que mesmo profissionais experientes podem ser influenciados por fatores claramente aleatórios durante o processo decisório.

2.3.5 Vieses cognitivos particularmente presentes da atividade processual: confirmação, retrospectiva, falácia da conjunção, impossibilidade de esquecimento de provas ilícitas e vieses de decisões sequenciais

Alguns tipos de vieses são particularmente identificados no campo jurídico, especialmente no processo de tomada de decisões judiciais.

Um primeiro exemplo consiste no *viés de confirmação* (“confirmation bias”). Por ele, se as pessoas têm uma compreensão prévia ou uma hipótese a respeito de um determinado assunto, elas tendem a selecionar as informações disponíveis, em determinado caso, que correspondem às compreensões prévias, ignorando as evidências em sentido contrário⁴⁰.

Um importante estudo a respeito desse particular viés foi conduzido pela Universidade de Stanford, em que participantes favoráveis ou contrários à pena de morte foram instruídos a ler determinados estudos sobre o tema, os quais apresentavam ambos os pontos de vista. Referido experimento demonstrou que os participantes deram mais credibilidade aos estudos atrelados às suas prévias convicções⁴¹.

39. ENGLISH, Birt; MUSSWEILER, Thomas. Sentencing Under Uncertainty: Anchoring Effects in the Courtroom. *Journal of Applied Social Psychology*, v. 31, n. 7, p. 1535-1551, 2001.

40. PEER, Eyal; GAMLIEL, Eyal. Heuristics and Biases in Judicial Decisions. *Court Review*, v. 49, n. 2, p. 114-119, 2013, p. 115.

41. LORD, Charles G.; ROSS, Lee; LEPPER, Mark R. Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, v. 37, n. 11, p. 2098-2109, 1979. Disponível em: https://www.researchgate.net/publication/232555483_Biased_Assimilation_and_Attitude_Polarization_The_Effects_of_Prior_Theories_on_Subsequently_Considered_Evidence. Acesso em 2 set. 2020.

Aplicado no campo processual, o viés de confirmação é capaz de afetar juízes no processo de avaliação das provas produzidas, descartando aquelas contrárias à compreensão ou hipótese previamente assentada.

Essa possibilidade foi identificada por Rassin, Eerland e Kuijpers, em estudo cujos participantes foram apresentados a um caso de homicídio, em que a vítima era uma psicóloga. A principal suspeita apresentada era a esposa de um dos pacientes, que supostamente teria matado a vítima por ciúmes. Os participantes foram então instruídos a avaliar vinte peças de informação e classificar a probabilidade de eles conduzirem à condenação ou absolvição da principal suspeita. Todavia, metade dos participantes foi apresentada à possibilidade de um outro suspeito: um antigo paciente da vítima, que a perseguia por um tempo considerável. Surpreendentemente, todos os participantes classificaram as peças de evidências de forma similar, indicando que a primeira suspeita seria culpada na mesma proporção. Segundo concluíram os pesquisadores, os participantes apenas levaram em consideração as evidências que reforçavam suas hipóteses iniciais⁴².

Um outro importante elemento consiste no *viés de retrospectiva* (“hindsight bias”). Por ele, quando as pessoas avaliam eventos ou resultados após o seu acontecimento, existe uma probabilidade maior de julgadores considerarem o fato mais previsível do que na época em que ocorreu.

Esse viés possui uma relação direta com casos de responsabilidade civil, especialmente quando se busca apurar situações relativas a condutas negligentes. Um caso ilustrativo, descrito por Peer e Gamliel, diz respeito à um processo movido em desfavor de um médico acusado de negligência por falhar na detecção de um pequeno tumor de um paciente, em um exame de radiografia preliminar. O médico foi condenado após outro radiologista – que examinou as radiografias após o tumor ter sido encontrado – testemunhar no sentido de que o tumor poderia ter sido identificado no exame preliminar. Obviamente, o segundo radiologista tinha, como elemento facilitador, a vantagem de saber que o tumor realmente existia ali, algo que o primeiro médico desconhecia⁴³.

Em terceiro lugar, também merece destaque a *falácia da conjunção* (“conjunction fallacy”), que diz respeito ao julgamento que as pessoas fazem a respeito da probabilidade de um evento ter ocorrido com base nos detalhes da sua descrição.

42. RASSIN, Eric; EERLAND, Anieta; KUIJPERS, Ilse. Let’s Find the Evidence: An Analogue Study of Confirmation Bias in Criminal Investigations. *Journal of Investigative Psychology and Offender Profiling*, v. 7, p. 231-246, 2010. Disponível em: https://www.researchgate.net/publication/230268983_Let’s_Find_the_Evidence_An_Analogue_Study_of_Confirmation_Bias_in_Criminal_Investigations. Acesso em 2 set. 2020.

43. PEER, Eyal; GAMLIEL, Eyal. Heuristics and Biases in Judicial Decisions. *Court Review*, v. 49, n. 2, p. 114-119, 2013, p. 115.

Pessoas acreditam que eventos descritos dessa forma são mais prováveis que aqueles narrados com menos detalhes.

Curiosamente, no campo da teoria estatística clássica, eventos menos detalhados, em realidade, podem ser desdobrados em outros eventos mais detalhados e, por isso mesmo, não podem ser considerados menos prováveis. Por exemplo, a probabilidade de um determinado indivíduo ser condenado “pela prática de um crime” (menos detalhes) não pode ser menor que as chances de ser condenado “pela prática do crime X”.

Esse viés foi identificado em estudo descrito por Fox e Birke, em que os participantes – advogados experientes – foram questionados sobre a possibilidade de um caso específico (Jones vs. Clinton) ter um desfecho outro que não uma sentença judicial. De forma surpreendente, o experimento revelou que a soma das probabilidades que os participantes assinalaram relativas a desfechos como a desistência do caso, sua inadmissibilidade ou celebração de acordo foi superior à probabilidade de o mérito do caso “não ser solucionado por decisão judicial”⁴⁴.

Um outro relevante viés diz respeito à *impossibilidade de se ignorar a prova excluída*. Em alguns casos judiciais, a ilicitude de uma determinada prova conduz à sua retirada dos autos, não sendo possível a sua utilização para fins de fundamentação da decisão a ser tomada. Estudos revelam, porém, que a probabilidade de a prova excluída ser levada em consideração, ainda que os jurados sejam instruídos a não o fazer, é considerável⁴⁵.

O viés de decisões sequenciais (“biased decisions in sequential ruling”) também merece destaque. Por ele, em sessões ou audiências com decisões sequenciais, os julgadores tendem a decidir em favor do *status quo* ao longo do tempo, muito embora fatores externos – como uma simples parada para almoço – afetem essa tendência.

Um estudo conduzido por Danziger sobre juízes em Israel demonstrou que 65% (sessenta e cinco por cento) das decisões eram favoráveis ao requerente no início de cada sessão, logo após o café da manhã, regredindo para até 10% (dez por cento) no final de cada sessão. Segundo os autores, as decisões seguidas, proferidas nas sessões, esgotavam os recursos mentais dos juízes, aumentando a chance de concessão de fiança em casos analisados após o intervalo⁴⁶.

44. FOX, Craig R.; BIRKE, Richard. Forecasting Trial Outcomes: Lawyers Assign Higher Probability to Possibilities That Are Described in Greater Detail. *Law and Human Behavior*, v. 26, n. 2, p. 159-173, 2002. Disponível em: https://www.researchgate.net/publication/11385112_Forecasting_Trial_Outcomes_Lawyers_Assign_Higher_Probability_to_Possibilities_That_Are_Described_in_Greater_Detail. Acesso em 2 set. 2020.

45. LIEBERMAN, Joel D.; ARNDT, Jamie. Understanding the Limits of Limiting Instructions: Understanding the limits of limiting instructions: Social Psychological Explanations for the Failures of Instructions to Disregard Pretrial Publicity and Other Inadmissible Evidence. *Psychology Public Policy and Law*, v. 6, p. 677-711, 2000.

46. DANZIGER, Shai; LEVAV, Jonathan; AVNAIM-PESSO, Liora. Extraneous factors in judicial decisions. *PNAS*, v. 108, n. 17, p. 6889-6892, 2011.

2.4 Como confrontar os vieses cognitivos no setor público?

Os vieses cognitivos, como exposto, são uma decorrência de processos heurísticos inconscientes de qualquer ser humano. No âmbito público, contudo, são muitas as preocupações quanto aos possíveis impactos em processos decisórios conduzidos por agentes públicos (administradores, juízes, membros do Ministério Público etc.). De fato, como visto, tais vieses podem comprometer não apenas o devido processo legal – em seu aspecto formal –, mas também a justiça da decisão proferida.

Uma das principais medidas indicadas para minimizar os efeitos deletérios dos vieses cognitivos consiste no incremento da educação sobre o assunto. Atualmente, a grande maioria dos pesquisadores do tema reconhece que o primeiro grande passo para eliminar os efeitos dos vieses implícitos reside no reconhecimento de que esse problema existe⁴⁷. A mera tomada de consciência (a “normalização”), embora seja algo relativamente simples, é capaz de, ao fomentar a reflexão pessoal, reduzir sensivelmente os vieses.

O problema, todavia, está na impossibilidade de agentes públicos fazerem uma análise individualizada de cada caso que se lhes apresenta, dedicando a atenção necessária para que a sua forma de pensar não se entregue integralmente aos processos heurísticos.

Em razão dessa impossibilidade, a tecnologia surge – a o lado da normalização e da regulação – como uma possível ferramenta auxiliar. Para tanto, é necessária uma análise acurada dos riscos e benefícios da sua utilização.

3. O USO DA INTELIGÊNCIA ARTIFICIAL PODE CONTRIBUIR PARA REFORÇAR OS VIESES COGNITIVOS?

Cada vez mais, decisões humanas são delegadas a ferramentas de inteligência artificial desenvolvidas a partir do uso de algoritmo, assim compreendido o conjunto finito de instruções que, seguidas, realizam uma tarefa específica⁴⁸.

Dentro ou fora das ferramentas de inteligência artificial, os algoritmos são marcados pelas seguintes características:

47. BIWER, Meagan. Implicit Bias in the Judiciary: Confronting the Problem through Normalization, *Ind. J. L. & Soc. Equal.*, v. 7, 2019, p. 268. Cf. também: RACHLINSKI, Jeffrey J.; JOHNSON, Sheri; WISTRICH, Andrew J.; GUTHRIE, Chris. Does unconscious racial bias affect trial judges? *Cornell Law Faculty Publications*, 2009. Disponível em: <https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1691&context=facpub>. Acesso em: 30 ago. 2020.
48. PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência artificial e Direito*. Curitiba: Alteridade, 2019, p. 71.

1) *input* fornecido externamente; 2) *output* quantificável produzido; 3) *definiteness*, sendo cada instrução clara e inequívoca; 4) *finiteness*, terminando após um número finito de etapas; 5) *effectiveness*, com cada instrução sendo básica o suficiente para ser viável e executada por uma pessoa⁴⁹.

Qualquer que seja o seu ambiente de aplicação, os algoritmos funcionam a partir de dados (*inputs*) que lhes são fornecidos, para, a através de uma sequência finita de regras, produzir um resultado (*output*), a exemplo de uma decisão administrativa ou judicial. No estágio atual da inovação tecnológica do Direito, essas regras são especialmente voltadas à classificação de casos, documentos, eventos ou pessoas, havendo um contínuo avanço para tarefas bastante sofisticadas.

Problemas podem surgir em todas as etapas da aplicação do algoritmo, sobretudo na definição das suas regras e no emprego de dados indevidos. Diante disso, são crescentes os trabalhos vocacionados a levantar preocupações quanto a três grandes problemas que podem decorrer do seu uso: a) o emprego de *data sets (inputs)* viciados; b) a opacidade dos algoritmos não programados; c) a discriminação que pode ser gerada por algoritmos de *machine learning*⁵⁰.

É possível, por exemplo, que uma ferramenta processual destinada a sugerir ao julgador uma determinada sanção aplicável a um condenado em processo criminal incorpore os mesmos vieses da pessoa a quem busca auxiliar, absorvendo, seja nas regras ou nos dados, os vícios já existentes.

Nessa linha, Dierle Nunes e Ana Luiza Pinto Coelho Marques exemplificam diversos casos de modelos enviesados, identificados em ferramentas de uso comum:

Também é possível verificar vieses algorítmicos no sistema de concessão de crédito europeu e norte-americano, na medida em que diversas companhias utilizam modelos de IA para análise do risco do empréstimo. Muitos desses modelos utilizam até mesmo dados das redes sociais do solicitante para o cálculo do *credit score*, baseando-se, assim, nas conexões sociais do indivíduo. Dessa forma, o resultado vincula-se diretamente ao grupo social no qual o solicitante está inserido. Corroborando tal fato, um relatório de 2007, apresentado pela *Federal Reserve* ao Congresso dos Estados Unidos, apontou que negros e hispânicos têm um *credit score* significativamente inferior ao de brancos e asiáticos.

Outros exemplos de modelos enviesados podem ser citados: um sistema de reconhecimento facial criado pela Google identificou pessoas negras como gorilas; o sistema de busca de contatos do aplicativo LinkedIn demonstrou uma preferência por

49. PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência artificial e Direito*. Curitiba: Alteridade, 2019, p. 71.

50. FERRARI, Isabela; BECKER, Daniel; WOLKART, Erik Navarro. Arbitrium ex machina: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos. *Revista dos Tribunais*, v. 995, p. 635-655, set./2018.

nomes de pessoas do sexo masculino; Tay, mecanismo de IA lançado pela Microsoft para interagir com usuários do Twitter, passou a reproduzir mensagens xenofóbicas, racistas e antisemitas; o aplicativo de *chat* SimSimi, que utiliza inteligência artificial para conversar com os usuários, foi suspenso no Brasil por reproduzir ameaças, palavrões e conversas de teor sexual.⁵¹

Muito embora os exemplos citados tenham ocorrido no campo das instituições privadas, as ferramentas de inteligência artificial também têm sido utilizadas no setor público, auxiliando o processo de tomadas de decisões.

É o que ocorre com o projeto VICTOR, apresentado pelo Supremo Tribunal Federal brasileiro aos jurisdicionados em 26 de setembro de 2018, por ocasião do II Congresso Internacional de Direito, Governo e Tecnologia. Cuida-se de ferramenta de inteligência artificial, fruto de parceria do tribunal com a Universidade de Brasília (UnB), que, no momento, está sendo utilizada na Corte para separação e classificação das peças do processo judicial e identificação dos principais temas de repercussão geral⁵².

No Brasil, além do projeto VICTOR, podem ser citados os robôs Alice, Sofia e Monica, que têm sido empregados pelo Tribunal de Contas da União para, respectivamente: a) examinar milhares de editais de licitação e atas de preços em busca de fraudes e irregularidades; b) analisar e sugerir aprimoramentos em relatórios internos; e c) acompanhar todas as compras públicas, inclusive as decorrentes de contratação direta. Em razão do sucesso de tais instrumentos, seu uso foi transportado para outros órgãos de controle, a exemplo dos Tribunais de Contas estaduais, Ministério Público, Polícia Federal e Controladoria Geral da União⁵³.

Em alguns casos, o uso da inteligência artificial não se resume à classificação processual, auxiliando diretamente o processo decisório. A título de exemplo, o *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) consiste em um mecanismo utilizado nos Estados Unidos para avaliar o risco de reincidência dos acusados no país, auxiliando os juízes a decidirem a respeito das prisões cautelares⁵⁴.

51. NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, v. 285, p. 421-447, nov./2018.

52. Cf. <http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=390818>. Acesso em: 10 ago. 2019.

53. FERRARI, Isabela; BECKER, Daniel; WOLKART, Erik Navarro. Arbitrium ex machina: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos. *Revista dos Tribunais*, v. 995, p. 635-655, set./2018.

54. NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, v. 285, p. 421-447, nov./2018.

O COMPAS já foi apontado como um sistema que produz resultados discriminatórios. Segundo artigo publicado por autores vinculados à ProPublica – instituição sem fins lucrativos sediada em Nova York que tem por missão a realização de jornalismo investigativo de interesse público –, o algoritmo utilizado tende a classificar erroneamente acusados negros como prováveis reincidentes, além de enquadrar, também de forma equivocada, acusados brancos como indivíduos com baixo risco de reincidência.

Demais disso, a empresa desenvolvedora do *software* – Northpointe – não disponibiliza ao público o algoritmo no qual se baseia o índice de reincidência do acusado, limitando-se a divulgar as perguntas feitas aos acusados, cujas respostas são utilizadas no cálculo. Consequentemente, o acusado não é capaz de conhecer as razões pelas quais possui um alto ou baixo indicador, tampouco a forma com que suas respostas afetam o resultado final⁵⁵.

Em 2017, ao julgar o caso *State v. Loomis*, a Suprema Corte de Wisconsin decidiu que a utilização do sistema COMPAS, a despeito da ausência de compreensão a respeito de seu funcionamento pelo acusado, não viola o devido processo legal, nem compromete o direito de acesso à informação⁵⁶.

Casos como esse revelam preocupações fundamentais, de ordem democrática: a necessidade de se assegurar o direito fundamental de acesso à informação a todos os cidadãos sujeitos a tais mecanismos; e o dever do Estado de prevenir e corrigir as distorções decorrentes do desenvolvimento, alimentação ou uso inadequado das ferramentas, evitando que as decisões produzidas sejam enviesadas e discriminatórias. Tais necessidades objetivam evitar a criação do que a matemática Cathy O’Neil denominou “Weapons of math destruction”⁵⁷.

55. Nessa linha: “[...] the algorithm is more likely to misclassify a black defendant as higher risk than a white defendant. Black defendants who do not recidivate were nearly twice as likely to be classified by COMPAS as higher risk compared to their white counterparts (45 percent vs. 23 percent). However, black defendants who scored higher did recidivate slightly more often than white defendants (63 percent vs. 59 percent). The test tended to make the opposite mistake with whites, meaning that it was more likely to wrongly predict that white people would not commit additional crimes if released compared to black defendants. COMPAS under-classified white reoffenders as low risk 70.5 percent more often than black reoffenders (48 percent vs. 28 percent). The likelihood ratio for white defendants was slightly higher 2.23 than for black defendants 1.61” (LARSON, Jeff; MATTU, Surya; LAUREN, Kirchner; ANGWINHOW, Julia. *We Analyzed the COMPAS Recidivism Algorithm*. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 10 ago. 2019.

56. *Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S.Ct. 2290 (2017). Cf. ISRANI, Ellora. *Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis*. Disponível em: <https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1>. Acesso em: 10 ago. 2019.

57. O’NEIL, Cathy. *Weapons of math destruction: how big data increases inequality and threatens democracy*. Nova York: Crown, 2016.

A essencialidade da promoção de um elevado grau de transparência algorítmica, de modo a possibilitar o conhecimento daquilo que determina o resultado pela ferramenta, foi objeto de recente reflexão pelo Parlamento Europeu, que, em resolução de 16 de fevereiro de 2017, tratou do seguinte princípio ético para orientar a regulação da robótica:

12. Realça o princípio da transparência, nomeadamente o facto de que deve ser sempre possível fundamentar qualquer decisão tomada com recurso a inteligência artificial que possa ter um impacto substancial sobre a vida de uma ou mais pessoas; considera que deve ser sempre possível reduzir a computação realizada por sistemas de IA a uma forma compreensível para os seres humanos; considera que os robôs avançados deveriam ser dotados de uma “caixa preta” com dados sobre todas as operações realizadas pela máquina, incluindo os passos da lógica que conduziu à formulação das suas decisões.⁵⁸

De fato, não se pode correr o risco de que as constituições e declarações de direitos humanos fiquem reféns das inovações tecnológicas, cujo funcionamento, não raramente, não é transparente – ou, ao menos, não é facilmente apreendido pelos cidadãos em geral.

4. O USO DA INTELIGÊNCIA ARTIFICIAL PODE CONTRIBUIR PARA EVITAR OS VIESES COGNITIVOS?

Se, por um lado, o uso de ferramentas de inteligência artificial pode automatizar e intensificar, de forma fria e ainda mais opaca, os vieses implícitos, estudos sugerem que a sua adequada aplicação pode ensejar o efeito diametralmente oposto.

Em razão dos vieses da heurística de representatividade, foi possível perceber que a valoração moral de dados como os antecedentes de um litigante, sua conduta e até mesmo a sua etnia podem influenciar o processo decisório do juiz. Cuida-se de uma consequência indesejada dos processos heurísticos. Seria possível programar uma aplicação para que não padeça desse tipo de preconceito?⁵⁹ Embora a resposta a esse questionamento dependa de um labor experimental ainda embrionário, essa medida nos parece possível.

58. Parlamento Europeu. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103(INL)). Disponível em: <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2017-0051+0+DOC+XML+V0//PT>. Acesso em: 10 ago. 2019. No Brasil, a Resolução nº 332/2020, do Conselho Nacional de Justiça, dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário.

59. FENOLL, Jordi Nieva. *Inteligencia artificial y proceso judicial*. Marcial Pons: Madrid, Barcelona, Buenos Aires, São Paulo, 2018, p. 48.

Quanto à representatividade, o que se observa é que alguns dados específicos de um determinado caso são, inconscientemente, apreendidos como algo mais chamativo àquele a quem foi atribuído o poder decisório. Ao menos em tese, um programador sensível às variáveis inconscientemente incorporadas nos processos decisórios é capaz de adotar as cautelas necessárias para que os dados objetivamente irrelevantes não influenciem a sugestão apresentada por meio de algoritmos de inteligência artificial. Mesmo os dados supostamente relevantes, muitas vezes, são superdimensionados no processo de tomada de decisão, razão pela qual seria possível cogitar, para as tarefas mais simples, a adoção de uma tabela de valores disponível às pessoas em geral.

A título de exemplo, a etnia de um determinado réu, embora seja um dado inconscientemente considerado por muitos julgadores – como demonstram os estudos anteriores –, é (ou ao menos deveria ser) algo absolutamente irrelevante em processos punitivos. Bastaria a um sistema automatizado de dosimetria da pena desconsiderar esse elemento ao formular uma sugestão ao órgão jurisdicional, diferentemente do que parece fazer o COMPAS.

De igual modo, ferramentas de inteligência artificial podem servir como forma de mitigar os riscos inerentes aos vieses decorrentes da heurística de disponibilidade. Isso porque, ao contrário dos seres humanos, a memória das máquinas é “linear”. Enquanto seres humanos correm o risco de valorar excessivamente a probabilidade de um evento em função da facilidade que tenham para recordá-lo, as máquinas não se submetem a uma “memória emotiva”.

Em outras palavras, como registra Fenoll, “uma aplicação de inteligência artificial poderia estar livre de vieses inspirados por variáveis tão emocionais”⁶⁰, a depender da forma com que se elabore o algoritmo. De forma exemplificativa, o desenvolvimento de premissas objetivas que possam ser diretamente associadas ao pressuposto do *periculum* das medidas cautelares é capaz de evitar decisões injustas, baseadas em presunções dissociadas da realidade fática.

Essa possibilidade parece ainda mais clara em relação ao viés de ancoragem e ajustamento.

De um lado, uma máquina tende sempre à ancoragem e ajuste, na medida em que sempre atuará de acordo com o que define o seu algoritmo. Se acaso o algoritmo for formulado de forma defeituosa, a máquina não poderá mudar de “opinião”, produzindo sempre o mesmo resultado. Sob esse aspecto, a ação humana é evidentemente superior à inteligência artificial.

Por outro lado, bem desenvolvido o seu algoritmo, cada dado introduzido na ferramenta será necessariamente considerado, sendo capaz de produzir resultados novos (novas “opiniões”), desprovidos do viés de ancoragem humano, que tende a

60. FENOLL, Jordi Nieva. *Inteligencia artificial y proceso judicial*. Marcial Pons: Madrid, Barcelona, Buenos Aires, São Paulo, 2018, p. 50.

dimensionar excessivamente um valor inicialmente sugerido ou desconsiderar variáveis relevantes, perpetuando uma ideia original (viés de confirmação). Consequentemente, é possível que a superação desde processo heurístico também passe pelo desenvolvimento de ferramentas de inteligência artificial⁶¹.

Para além disso, evidentemente, por serem desprovidas de emoções, as máquinas são capazes de abstrair por completo elementos que, conquanto aparentemente irrelevantes, influenciam o processo decisório: as vestimentas dos advogados, a fome, o tom de voz do interlocutor etc.

5. CONCLUSÃO

Ao final do exposto, apresentam-se as seguintes conclusões, sem prejuízo de outras ilações realizadas ao longo do texto:

1. Há muito, o dogma da ação racional, elemento central da economia clássica, tem sido confrontado por experimentos no campo da psicologia cognitiva e da economia comportamental (“behavioral economics”), especialmente a partir da pesquisa seminal realizada na década de 1970, por Amos Tversky e Daniel Kahneman. Os experimentos conduzidos por tais autores foram capazes de revelar que, durante processos decisórios, muitas pessoas se valem de um limitado número de princípios heurísticos automatizados,

2. Esses princípios heurísticos consistem em atalhos mentais utilizados na redução da complexidade das tarefas de conhecimento de situações e tomadas de decisões. Além de adaptativas, as heurísticas são inegavelmente úteis, contribuindo para uma maior eficiência nas decisões diárias. Apesar disso, elas podem resultar em erros severos e sistemáticos, conduzindo a falácias, vieses e ilusões.

3. Muitos juízes ainda desconhecem os vieses cognitivos implícitos e, em muitos casos – a exemplo das questões relativas a preconceitos raciais –, tratam o assunto como tabu. No atual cenário das questões éticas relativas à prestação jurisdicional, julgadores continuam livres para atuar de forma despreocupada quanto aos vieses implícitos que habitam o inconsciente. Esses vieses podem trazer consequências diversas, que variam desde a má apreciação das provas de um caso até associações estereotipadas.

4. Em um cenário marcado pelo intenso desenvolvimento de mecanismos de automação decisória, é possível que uma ferramenta processual destinada a sugerir ao julgador uma determinada decisão incorpore os mesmos vieses de quem busca auxiliar, absorvendo, seja nas regras ou nos dados, os vícios já existentes.

61. FENOLL, Jordi Nieva. *Inteligencia artificial y proceso judicial*. Marcial Pons: Madrid, Barcelona, Buenos Aires, São Paulo, 2018, p. 51.

5. Se, por um lado, o uso de ferramentas de inteligência artificial pode automatizar e intensificar, de forma fria e ainda mais opaca, os vieses implícitos, estudos sugerem que a sua adequada aplicação pode ensejar o efeito diametralmente oposto. Ao menos em tese, um programador sensível às variáveis inconscientemente incorporadas nos processos decisórios é capaz de adotar as cautelas necessárias para que os dados objetivamente irrelevantes não influenciem a sugestão apresentada por meio de algoritmos de inteligência artificial. Mesmo os dados supostamente relevantes, muitas vezes, são superdimensionados no processo de tomada de decisão, razão pela qual seria possível cogitar, para as tarefas mais simples, a adoção de uma tabela de valores disponível às pessoas em geral, mitigando-se os vieses da heurística de representatividade.

6. De igual modo, ferramentas de inteligência artificial podem servir como forma de mitigar os riscos inerentes aos vieses decorrentes da heurística de disponibilidade. Isso porque, ao contrário dos seres humanos, a memória das máquinas é “linear”. Enquanto seres humanos correm o risco de valorar excessivamente a probabilidade de um evento em função da facilidade que tenham para recordá-lo, as máquinas não se submetem a uma “memória emotiva”.

7. Essa possibilidade parece ainda mais clara em relação ao viés de ancoragem e ajustamento. De um lado, uma máquina tende sempre à ancoragem e ajuste, na medida em que sempre atuará de acordo com o que define o seu algoritmo. Por outro lado, bem desenvolvido o seu algoritmo, cada dado introduzido na ferramenta será necessariamente considerado, sendo capaz de produzir resultados novos (novas “opiniões”), desprovidos do viés de ancoragem humano, que tende a dimensionar excessivamente um valor inicialmente sugerido ou desconsiderar variáveis relevantes, perpetuando uma ideia original (viés de confirmação). Consequentemente, é possível que a superação deste processo heurístico também passe pelo desenvolvimento de ferramentas de inteligência artificial.

REFERÊNCIAS

- BIWER, Meagan. Implicit Bias in the Judiciary: Confronting the Problem through Normalization, *Ind. J. L. & Soc. Equal*, v. 7, 2019.
- BURCH, Traci. Skin Color and the Criminal Justice System: Beyond Black-White Disparities in Sentencing. *Journal of Empirical Legal Studies*, v. 12, p. 395-420, 2015.
- CARROLL, John S. The effect of imagining an event on expectations for the event: an interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology*, v. 14, n. 1, p. 88-96, 1978.
- DANZIGER, Shai; LEVAV, Jonathan; AVNAIM-PESSO, Liora. Extraneous factors in judicial decisions. *PNAS*, v. 108, n. 17, p. 6889-6892, 2011.
- DYSON, Freeman. How to Dispel Your Illusions. *New York Review of Books*, dez/2011. Disponível em: <http://www.nybooks.com/articles/archives/2011/dec/22/how-dispel-your-illusions/>. Acesso em: 23 ago. 2020.

- EBERHARDT, Jennifer L; DAVIES, Paul G.; PURDIE-VAUGHNS, Valerie J.; JOHNSON, Sheri Lynn. Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes. *Psychological Science*, v. 17, p. 383-386, 2006.
- ENGLISH, Birt; MUSSWEILER, Thomas. Sentencing Under Uncertainty: Anchoring Effects in the Courtroom. *Journal of Applied Social Psychology*, v. 31, n. 7, p. 1535-1551, 2001.
- ENGLISH, Birt; MUSSWEILER, Thomas; STRACK, Fritz. Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making. *Personality and Social Psychology Bulletin*, v. 32, p. 188-200, 2006.
- FENOLL, Jordi Nieva. *Inteligencia artificial y proceso judicial*. Marcial Pons: Madrid, Barcelona, Buenos Aires, São Paulo, 2018.
- FENOLL, Jordi Nieva. Transfondo psicológico de la independencia judicial. In: FENOLL, Jordi Nieva; OTEIZA, Eduardo (dirs.). *La independência judicial: um constante asedio*. Marcial Pons: Madri, Barcelona, Buenos Aires, São Paulo, 2019, p. 24.
- FERRARI, Isabela; BECKER, Daniel; WOLKART, Erik Navarro. Arbitrium ex machina: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos. *Revista dos Tribunais*, v. 995, p. 635-655, set./2018.
- FOX, Craig R.; BIRKE, Richard. Forecasting Trial Outcomes: Lawyers Assign Higher Probability to Possibilities That Are Described in Greater Detail. *Law and Human Behavior*, v. 26, n. 2, p. 159-173, 2002. Disponível em: https://www.researchgate.net/publication/11385112_Forecasting_Trial_Outcomes_Lawyers_Assign_Higher_Probability_to_Possibilities_That_Are_Described_in_Greater_Detail. Acesso em 2 set. 2020.
- GOFF, Phillip Atiba; JACKSON, Matthew Christian; DI LEONE, Brooke Allison Lewis; CULOTTA, Carmen Marie; DITOMASSO, Natalie Ann. The Essence of Innocence: Consequences of Dehumanizing Black Children. *Journal of Personality and Social Psychology*, v. 106, n. 4, p. 526-545, 2014. Disponível em: <https://www.apa.org/pubs/journals/releases/psp-a0035663.pdf>. Acesso em: 30 ago. 2020.
- JONES, Craig. The Troubling New Science of Legal Persuasion: Heuristics and Biases in Judicial Decision-Making. *Advocates Quaterly*, v. 41, p. 49-122, 2013.
- JONES, Craig; RANKIN, Micah B. E. Justice as a Rounding Error? Evidence of Subconscious Bias in Second-Degree Murder Sentences in Canada. *Osgoode Digital Commons*, v. 10, n. 81, 2014.
- JONES, Craig E. The Troubling New Science of Legal Persuasion: Heuristics and Biases in Judicial Decision-Making. *Advocates' Quarterly*, v. 41, n. 1, 2013.
- KAHNEMAN, Daniel. *Rápido e Devagar*. Tradução: Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012.
- KAHNEMAN, Daniel. TVERSKY, Amos. *On the study of statistical intuitions*. *Cognition*, 1982, p. 123.
- LARSON, Jeff; MATTU, Surya; LAUREN, Kirchner; ANGINHOW, Julia. *We Analyzed the COMPAS Recidivism Algorithm*. Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 10 ago. 2019.
- LIEBERMAN, Joel D.; ARNDT, Jamie. Understanding the Limits of Limiting Instructions: Understanding the limits of limiting instructions: Social Psychological Explanations for the

- Failures of Instructions to Disregard Pretrial Publicity and Other Inadmissible Evidence. *Psychology Public Policy and Law*, v. 6, p. 677-711, 2000.
- LORD, Charles G.; ROSS, Lee; LEPPER, Mark R. Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, v. 37, n. 11, p. 2098-2109, 1979. Disponível em: https://www.researchgate.net/publication/232555483_Biased_Assimilation_and_Attitude_Polarization_The_Effects_of_Prior_Theories_on_Subsequently_Considered_Evidence. Acesso em 2 set. 2020.
- NEGOWETTI, Nicole. E. Judicial Decisionmaking, Empathy and the Limits of Perception. *Akron Law Review*, v. 47, 2014.
- NUNES, Dierle; MARQUES, Ana Luiza Pinto Coelho. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, v. 285, p. 421-447, nov./2018.
- NUNES, Dierle; LUD, Nathanael; PEDRON, Flávio. *Desconfiando da imparcialidade dos sujeitos processuais*. Salvador: Juspodivm, 2018.
- OAKES, Anne Richardson; DAVIES, Haydn. Process, Outcomes and the Invention of Tradition: The Growing Importance of the Appearance of Judicial Neutrality. *Santa Clara Law Review*, v. 51, n. 2, 2011.
- O'NEIL, Cathy. *Weapons of math destruction: how big data increases inequality and threatens democracy*. Nova York: Crown, 2016.
- PARKS, Gregory S.; DAVIS, Andre M. Confronting Implicit Bias: An Imperative for Judges in Capital Prosecutions. *Human Rights Magazine*, v. 42, n. 2, 2016. Disponível em: https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/2016-17-vol-42/vol--42--no--2---the-death-penalty--how-far-have-we-come-/confronting-implicit-bias--an-imperative-for-judges-in-capital-p/. Acesso em: 30 ago. 2020.
- PEER, Eyal; GAMLIEL, Eyal. Heuristics and Biases in Judicial Decisions. *Court Review*, v. 49, n. 2, p. 114-119, 2013.
- PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência artificial e Direito*. Curitiba: Alteridade, 2019.
- PEREIRA, Marcos Emanuel; ÁLVARO, José Luis; OLIVEIRA, Andréia C. Oliveira; DANTAS, Gilcimar. Estereótipos e essencialização de brancos e negros: um estudo comparativo. *Psicologia & Sociedade*, v. 1, p. 144-153, 2011.
- RACHLINSKI, Jeffrey J.; JOHNSON, Sheri; WISTRICH, Andrew J.; GUTHRIE, Chris. Does unconscious racial bias affect trial judges? *Cornell Law Faculty Publications*, 2009. Disponível em: <https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1691&context=facpub>. Acesso em: 30 ago. 2020.
- RASSIN, Eric; EERLAND, Anieta; KUIJPERS, Ilse. Let's Find the Evidence: An Analogue Study of Confirmation Bias in Criminal Investigations. *Journal of Investigative Psychology and Offender Profiling*, v. 7, p. 231-246, 2010. Disponível em: https://www.researchgate.net/publication/230268983_Let's_Find_the_Evidence_An_Analogue_Study_of_Confirmation_Bias_in_Criminal_Investigations. Acesso em 2 set. 2020.
- ROOTH, Dan-Olof. Implicit Discrimination in Hiring: Real World Evidence. *Institute for the Study of Labor (IZA)*, 2007. Disponível em: <http://ftp.iza.org/dp2764.pdf>. Acesso em 30 ago. 2020.

- STEIN, Christopher; DOUIN, Michelle. *Cognitive Bias in the Courtroom: Combating the Anchoring Effect in Criminal Sentencing*. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2991611. Acesso em: 31 ago. 2020.
- SUNSTEIN, Cass R. *Behavioral Law and Economics*. Cambridge: Cambridge University Press, 2000.
- TVERSY, Amos; KAHNEMAN, Daniel. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, v. 2, n. 5, p. 207-232, 1973.
- TVERSY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. *Science*, v. 185, p. 1.124-1.131, 1974.
- VITORELLI, Edilson. *Processo civil estrutural: teoria e prática*. Salvador: Juspodivm, 2020.
- WALDFOGEL, Joel; AYERS, Ian. A Market Test for Race Discrimination in Bail Setting. *Stanford Law Review*, v. 46, p. 987-1048, 1994.