THE COLUMBIA
SCIENCE & TECHNOLOGY
LAW REVIEW

VOLUME XXVI STLR.ORG FALL 2024

ARTICLE

AN ANATOMY OF ALGORITHM AVERSION

Cass R. Sunstein[*] and Jared H. Gaffe[†]

People are said to show "algorithm aversion" when they prefer human forecasters or decision-makers to algorithms, even though algorithms generally outperform people (in forecasting accuracy and/or optimal decision-making in furtherance of a specified goal). Algorithm aversion also has "softer" forms, as when people prefer human forecasters or decision-makers to algorithms in the abstract, without having clear evidence about comparative performance. Algorithm aversion has strong implications for policy and law; it suggests that those who seek to use algorithms, such as officials in federal agencies, might face serious public resistance. Algorithm aversion is a product of diverse mechanisms, including (1) a desire for agency; (2) a negative moral or emotional reaction to judgment by algorithms; (3) a belief that certain human experts have unique knowledge, unlikely to be held or used by algorithms; (4) ignorance about why algorithms perform well; and (5) asymmetrical forgiveness, or a larger negative reaction to algorithmic error than to human error. An understanding of the various mechanisms provides some clues about how to overcome algorithm aversion, and also of its boundary conditions. These clues bear on the numerous decisions in law and policy, including those of federal agencies (such as the Department of Homeland Security and the Internal Revenue Service) and those involved in the criminal justice system (such as those thinking about using algorithms for bail decisions).

[*] Robert Walmsley University Professor, Harvard University. We are grateful to the Harvard Initiative on Artificial Intelligence and the Law for valuable support.

[†] Harvard Law School, Class of 2026.

I. PUZZLES

It is widely said that people show "algorithm aversion."[1] If so, what is it, exactly, to which they are averse? For those who are interested in promoting the use of algorithms in policy and law—say, in the criminal justice system, in the area of tax policy,[2] or in the domain of immigration or refugee status[3]—algorithm aversion creates serious challenges.

The word "algorithm" is often taken to refer to a precise list of instructions that conduct specified actions step-by-step in either hardware- or software-based routines.[4] In common parlance, the term is usually reserved for sets of instructions

[1] *See* Ibrahim Filiz et al., *The Extent of Algorithm Aversion in Decision-Making Situations with Varying Gravity*, PLOS ONE, Feb. 21, 2023, at 1, 2-5, https://pubmed.ncbi.nlm.nih.gov/36809526/ [https://perma.cc/SYN2-9ZMT]; *see also* Cass R. Sunstein, *The Use of Algorithms in Society*, REV. OF AUSTRIAN ECON., May 4, 2023, at 11-13, https://csgs.kcl.ac.uk/wp-content/uploads/2023/05/s11138-023-00625-z.pdf [https://perma.cc/L8XD-AQCM] (discussing algorithm aversion and places where it commonly appears in society).

[2] *See* Hadi Elzayn et al., *Measuring and Mitigating Racial Disparities in Tax Audits* 33-34, 40 (unpublished working paper) (Jan. 30, 2023), https://dho.stanford.edu/wp-content/uploads/IRS_Disparities.pdf [https://perma.cc/52D7-PK65].

[3] *See* Niamh Kinchin, *Technology, Displaced? The Risks and Potential of Artificial Intelligence for Fair, Effective, and Efficient Refugee Status Determination*, 37 L. CONTEXT 45, 49-58 (2021).

[4] Sunstein, *supra* note 1, at 4; Alexander S. Gillis, *What Is an Algorithm?*, TECHTARGET (July 2024), https://www.techtarget.com/whatis/definition/algorithm [https://perma.cc/H8NJ-KNAR].

or calculations conducted by computers, and often refers to mechanisms involving machine learning and/or artificial intelligence.[5] An algorithm (1) takes a set of inputs, (2) conducts some set of computations and/or prioritizations, and (3) generates an output that may consist of predicted outcomes, probability assessments, synthesized analysis, summary information, or recommendations.[6] Why, it might be asked, would people be averse to that? As we shall soon see, the answer might lie in skepticism about certain technologies.

While familiar decision-making processes that neither require nor involve any technology or computation may meet a standard definition of "algorithm,"[7] these processes are not commonly associated with the term. For example, consider the following procedure for deciding whether to travel by taxi or public transportation: *take public transportation unless the Google Maps projected travel time via taxi is more than 20 minutes shorter than via public transportation*. In a sense, that is an algorithm. The commuter's thought process is a "procedure used for solving a problem or performing a computation," and almost surely a "set of rules to be followed in calculations or other problem-solving operations."[8] Still, the ordinary commuter likely does not think of a daily transportation decision as an "algorithm."[9]

People are often said to show "algorithm aversion" when (1) they prefer human forecasters or decision-makers to algorithms, even though (2) algorithms generally outperform people in the general domain or in a specific task (in forecasting accuracy and/or optimal decision-making in furtherance of an identifiable goal, e.g., predicting the likelihood that criminal defendants will flee the jurisdiction, or assessing the presence of heart disease or cancer).[10] In such cases, algorithm aversion appears to be a serious mistake.[11] Why would people be averse to a more accurate means of answering factual questions? Why would people reject the use of an algorithm that would (for example) save a large number of lives?[12]

[5] Sunstein, *supra* note 1, at 4.

[6] *See* Sunstein, *supra* note 1, at 4; *Algorithm*, NAT'L LIBR. MED. (May 25, 2022, 12:12 PM), https://www.nnlm.gov/guides/data-glossary/algorithm [https://perma.cc/HSR9-H4HM]; Gillis, *supra* note 4.

[7] *See* Sunstein, *supra* note 1, at 4 ("According to a standard definition, an algorithm is a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer. According to another, an algorithm is a procedure used for solving a problem or performing a computation.").

[8] *Id.*

[9] See DANIEL KAHNEMAN ET AL., NOISE 128-36 (2021), for relevant discussion regarding common human views of everyday occurrences, namely that they tend to view reality as a hindsight-driven set of inevitable causal links rather than considering all alternative scenarios given different inputs.

[10] *See* Filiz et al., *supra* note 1, at 3-4 (offering a list of definitions of algorithm aversion).

[11] Jens Ludwig et al., *The Unreasonable Effectiveness of Algorithms* 17-18 (Nat'l Bureau of Econ. Rsch., Working Paper No. 32125, 2024), https://www.nber.org/papers/w32125 [https://perma.cc/UJ7K-TWQU]. We say "appears to be" because there might be a specific (good) reason to favor a human judge in the particular case.

[12] See *id.* at 14-15 for valuable, brisk evidence.

Algorithm aversion is sometimes taken to occur when (1) people prefer human forecasters or decision-makers over algorithms even though (2) it is *unknown* whether algorithms outperform people (in forecasting accuracy and/or optimal decision-making in furtherance of a specified goal).[13] In such cases, algorithm aversion might or might not be a mistake. Finally, algorithm aversion might be taken to occur when (1) people prefer human forecasters or decision-makers over algorithms and (2) people generally outperform algorithms (in forecasting accuracy and/or in optimal decision-making in furtherance of a specified goal).[14] In such cases, algorithm aversion does not seem to be a mistake; people appear to be right to be averse to algorithms.[15]

Whatever its precise form, algorithm aversion has serious implications for public and private institutions and for policy and law. In many domains, agencies and courts are using algorithms, or are likely to use them in the future.[16] These institutions might be expected to run into serious resistance if they do so. If algorithms are more accurate than people in certain cases, and if lower accuracy leads to negative—sometimes severely so—real-world outcomes,[17] understanding algorithm aversion may offer important clues about how to overcome that aversion and thus prove helpful in improving those outcomes via greater algorithm adoption. Indeed, obtaining an understanding of algorithm aversion is our principal goal here.

In other cases, understanding why people are reluctant to use algorithms may help us to see what people seek in decision-making more broadly and thus help to

[13] *See generally* Filiz et al., *supra* note 1, at 3-4 (listing definitions of algorithm aversion, including aversion when the algorithm's efficacy is unknown); Sunstein, *supra* note 1, at 11-13 (suggesting reasons for algorithm aversion); Hasan Mahmud et al., *What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion*, 175 TECH. FORECASTING & SOC. CHANGE, Feb. 2022, at 1, 11-12, (suggesting that algorithm aversion may occur "even when there is no clear indication about the superiority of human decisions over algorithmic decisions").

[14] Mahmud et al., *supra* note 13, at 126-27.

[15] *See* Elzayn et al., *supra* note 2, at 21-26 (showing racial discrimination by an algorithm). The phrase "appear to be" is necessary because in the specific case, a human judge might be inferior to an algorithm.

[16] DAVID F. ENGSTROM ET AL., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 6-12 (2020); Cary Coglianese & Lavi Ben Dor, *AI in Adjudication and Administration*, 86 BROOK. L. REV. 791, 798-817 (2021); David F. Engstrom & Daniel E. Ho, *Artificially Intelligent Government: A Review and Agenda*, *in* RESEARCH HANDBOOK IN BIG DATA LAW 57, 59-61 (Roland Vogl ed., 2021); Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1160-67 (2017); s*ee e.g.*, Robert J. Kovacev, *Rise of the tax machines: IRS algorithms are coming for you*, THE HILL (Feb. 19, 2023, 12:00 PM), https://thehill.com/opinion/finance/3864905-rise-of-the-tax-machines-irs-algorithms-are-coming-for-you/ [https://perma.cc/ZZK5-2AX7] (written based on the premise that the IRS is currently using algorithms to trigger tax audits); OFFICE OF THE INSPECTOR GEN., SOC. SEC. ADMIN., THE SOCIAL SECURITY ADMINISTRATION'S USE OF INSIGHT SOFTWARE TO IDENTIFY POTENTIAL ANOMALIES IN HEARING DECISIONS 1 (2019) (clearly stating that the SSA uses algorithms in making some of its decisions); *see generally* U.S. DEP'T OF HOMELAND SEC., HIVE: A NOVEL ALGORITHMIC FRAMEWORK FOR STANDOFF CONCEALED THREAT DETECTION 1-5 (discussing an algorithm-based security system that was developed by an MIT lab funded by the Department of Homeland Security Science and Technology Directorate).

[17] *See* Sunstein, *supra* note 1, at 4-5.

identify the costs and benefits of algorithmic decision-making, which may not be limited to the accuracy of the relevant decisions. Most broadly, suppose that a public institution—say, the Department of Homeland Security—is seeking to rely on algorithms to make important decisions, or to assist in their making.[18] Will the public accept that decision? Or suppose that a company is shifting to reliance on algorithms for hiring and promotion. Will employees accept that decision? The answers to these questions have implications as well for reliance on generative AI and Large Language Models, though the underlying considerations may not be the same. Understanding the drivers of algorithm aversion may help address these issues for the private as well as the public sector.

A frequent finding, on which we elaborate below, is that in important cases, many people would indeed prefer to base their decisions or forecasts on advice from other human beings, or on their own judgments, rather than on decisions or forecasts from algorithms. In many of these cases, algorithmic performance is quantifiable; over time, results will show that an algorithm forecasted with greater/less accuracy or made decisions that promoted/did not promote an identifiable outcome. In some cases, the gains from the use of algorithms are truly extraordinary,[19] which means that algorithm aversion might be a serious problem for policymakers and those involved in the criminal justice system. A variety of reasons for the frequently superior performance of algorithms have been identified,[20] including algorithms' ability to reduce or eliminate the effects of (cognitive) biases (such as availability bias)[21] and noise in human judgment[22] and remove the potential for basic human errors (e.g., lack of information, miscalculation, reasoning errors, poor recall, typos, and so forth). Although people may do better than algorithms in important cases,[23] algorithm aversion is not limited to situations in which algorithmic performance is especially weak or where the factors that drive algorithmic outperformance do not apply. Algorithm aversion appears to affect decision-making in a variety of situations, including those in which algorithms do better than people.[24]

[18] *See* Open Funding Opportunities, U.S. DEP'T OF HOMELAND SEC., https://oip.dhs.gov/baa/public/funding-page?status=open [https://perma.cc/ZX23-GPWK] (last visited Sep. 6, 2024); *Artificial Intelligence at DHS*, U.S. DEP'T OF HOMELAND SEC., https://www.dhs.gov/ai [perma.cc/ZX23-GPWK] (last visited Sep. 6, 2024).

[19] *See* Ludwig et al., *supra* note 11, at 9-13.

[20] *See, e.g.*, *id.* at 17-18 (identifying several flaws that exist in human reasoning and not in algorithmic reasoning).

[21] Cass R. Sunstein, *Algorithms, Correcting Biases*, 86 SOC. RSCH. 499, 502-504 (2019); s*ee also* Ludwig et al., *supra* note 11, at 17 (discussing reliance on heuristics and bias in human judgment).

[22] Sunstein, *supra* note 1, at 10-11.

[23] *See, e.g.*, Kahneman et al., *supra* note 9, at 279-95 (discussing circumstances in which human decision-makers are equivalent or superior to algorithmic decision-makers).

[24] *See* Filiz et al., *supra* note 1, at 3-4; Timothy DeStefano et al., *Why providing humans with interpretable algorithms may, counterintuitively, lead to lower decision-making performance 26-28* (MIT Sloan Sch. Mgmt. Working Paper, Paper No. 6796, 2020), https://pdfs.semanticscholar.org/c7e7/7a4255809e4e43597ebbeee9be49c2ade7ca.pdf [https://perma.cc/N6GY-46BZ].

The opposite of algorithm aversion, the propensity or openness to choosing algorithms over human judgment, has been dubbed "algorithm appreciation."[25] We can find cases in which algorithm appreciation is generally sensible, cases in which we do not know whether it is sensible or not, and cases in which it seems to be a mistake. Relatively little research has focused on algorithm appreciation, but the topic is attracting growing attention.[26] Algorithm appreciation raises its own questions (it might be a mistake, or it might not be),[27] and while we will have a few things to say about it, we will largely bracket those questions here.

II. Diverse Mechanisms: An Overview

Algorithm aversion is driven by a wide variety of factors. At a general level, the drivers of algorithm aversion can be sorted into three categories: (1) internal factors related to a person's intrinsic needs and notion of self; (2) general factors related to a person's skepticism and comfort level with algorithms; and (3) task-specific factors related to perceptions about the ability of an algorithm to handle the issue at hand.

Here is an overview:

Table 1. Examples of sources of algorithm aversion described or found in existing literature

Category	Driver of Aversion	Examples
Internal Factors	**Agency and Control**	• Self-navigation instead of using a GPS or app • Vacation planning • Picking individual stocks rather than ETFs or robo-advising

[25] Hasan Mahmud et al., *Decoding Algorithm Appreciation: Unveiling the Impact of Familiarity with Algorithms, Tasks, and Algorithm Performance*, 179 Decision Support Sys., Apr. 2024, at 1, 1.

[26] *See, e.g.*, *id.* at 8-9 (discussing one example of such research); Melissa Saragih et al., *The Effect of Past Algorithmic Performance and Decision Significance on Algorithmic Advice Acceptance*, 38 Int'l J. Hum.-Comput. Interaction 1228, 1229-30, 1233-36 (2022) (discussing several past examples of research on algorithm appreciation, outlining a new piece of research, and suggesting further directions for future research); Jennifer M. Logg et al., *Algorithm appreciation: People prefer algorithmic to human judgment*, 151 Organizational Behav. & Hum. Decision Processes 90, 92 (2019) (discussing one example of current research on algorithm appreciation); Esther Kaufmann et al., *Task-Specific Algorithm Advice Acceptance: A Review and Directions for Future Research*, 7 Data & Info. Mgmt., Sept. 2023, at 1, 11-12 (discussing one example of research and suggesting future directions for other research).

[27] *See* Cass R. Sunstein & Lucia Reisch, *On Liking Algorithms*, Env't & Res. Econ. (forthcoming 2024) (manuscript at 7-8), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4787640 [https://perma.cc/N3W7-UZX9].

Category	Driver of Aversion	Examples
	Moral or Emotional Qualms	• Bail and release conditions[28] • Medical treatment[29]
	Unique Skills	• Skilled workers' reluctance to defer to algorithms[30]
	Human Failure, Algorithmic Failure	• Doctors' misdiagnoses • Poor legal and investment advice
	Loss Aversion, Risk Aversion	• Bond trader being less likely to use an algorithm that projects losses than returns • Employee deferring to algorithmic projections to avoid personal consequences of inaccuracy
	Confirmation Bias and Status Quo Bias	• Ride-sharing drivers' location choices[31]
	Herding and Conformity	• Ride-sharing drivers' location choices[32]
General Comfort and Skepticism	**Low Levels of Comfort with Innovation and Technology**	• Robo-advisers vs. human investment advisers
	Lack of Understanding and Trust	• Joke recommendation[33] • Social media algorithms
Task-Specific Confidence and	**Perceived Suitability for Algorithmic Decision-Making**	• Finding a romantic partner[34] • Naming children • Choosing a summer camp

[28] Jon Kleinberg et al., *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237, 239 (2017).

[29] Sendhil Mullainathan & Ziad Obermeyer, *Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*, 137 Q.J. ECON. 679, 680-83 (2022).

[30] Ryan T. Allen et al., *Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion*, 33 ORG. SCI. 149, 152-53 (2022).

[31] Meng Liu et al., *Algorithm Aversion: Evidence from Ridesharing Drivers*, MGMT. SCI., OCT. 3, 2023, at 1, 1-2.

[32] *Id.*

[33] Michael Yeomans et al., *Making Sense of Recommendations*, 32 J. BEHAV. DECISION MAKING 403, 404 (2019).

[34] Samantha Joel et al., *Is Romantic Desire Predictable? Machine Learning Applied to Initial Romantic Attraction*, 28 PSYCH. SCI. 1478, 1479 (2017).

Category	Driver of Aversion	Examples
Perceived Suitability	**High Levels of Confidence in Human Decision-Maker**	• Doctors' diagnoses[35] • Coaches' and scouts' lineup decisions • Judges' bail determinations[36]
	Ignored Factors and Unintended Consequences	• Employee promotions and team fit

Table 2. Overview of literature describing algorithm aversion

Category	Driver of Aversion	References in Existing Academic Literature
Internal Factors	**Agency and Control**	• Cass R. Sunstein, *The Use of Algorithms in Society*, Rev. of Austrian Econ., May 4, 2023. • Roy Shoval et al., *Choosing to Choose or Not*, 17 Judgment & Decision Making 768 (2022). • Sebastian Bobadilla-Suarez et al., *The Intrinsic Value of Choice: The Propensity to Under-Delegate in the Face of Potential Gains and Losses*, 54 J. Risk & Uncertainty 187 (2017). • Hasan Mahmud et al., *What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion*, 175 Tech. Forecasting & Soc. Change, Feb. 2022, at 1. • Cass R. Sunstein, *Choosing Not To Choose*, 64 Duke L.J. 1, 9 (2014).
	Moral or Emotional Qualms	• Daniel Kahneman, Thinking, Fast and Slow (2011) • Jon Kleinberg et al., *Human Decisions and Machine Predictions*, 133 Q.J. Econ. 237 (2017) • Sendhil Mullainathan & Ziad Obermeyer, *Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*, 137 Q.J. Econ. 679 (2022). • Ibrahim Filiz et al., *The Extent of Algorithm Aversion in Decision-Making Situations with Varying Gravity*, PLoS ONE, Feb. 21, 2023, at 1.

[35] *See* Sunstein & Reisch, *supra* note 27.

[36] *See id.*

Category	Driver of Aversion	References in Existing Academic Literature
	Unique Skills	• Nicolas Epley & Thomas Gilovich, *The Mechanics of Motivated Reasoning*, 30 J. Econ. Persps. 133 (2016). • Ryan T. Allen et al., *Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion*, 33 Org. Sci. 149 (2022). • Hasan Mahmud et al., *What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion*, 175 Tech. Forecasting & Soc. Change, Feb. 2022, at 1. • Nicole Tsz Yeung Liu et al., *Is algorithm aversion WEIRD? A cross-country comparison of individual-differences and algorithm aversion*, J. Retailing & Consumer Services, May 2023, at 1. • Noah Castelo, *Perceived corruption reduces algorithm aversion*, 34 J. Consumer Psych. 326 (2023).
	Human Failure, Algorithmic Failure	• Berkeley J. Dietvorst et al., *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. Experimental Psych. 114 (2015). • Alvaro Chacon et al., *A longitudinal approach for understanding algorithm use*, J. Behav. Decision Making, Oct. 2022 at 1.
	Loss Aversion, Risk Aversion	• Hasan Mahmud et al., *What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion*, 175 Tech. Forecasting & Soc. Change, Feb. 2022, at 1. • Inga Toma et al., *Impact of Loss and Gain Forecasting on the Behavior of Pricing Decision-making*, 6 Int'l J. Data Sci. & Analysis 12 (2020).
	Confirmation Bias and Status Quo Bias	• Meng Liu et al., *Algorithm Aversion: Evidence from Ridesharing Drivers*, Mgmt. Sci., Oct. 3, 2023, at 1. • Nicolas Epley & Thomas Gilovich, *The Mechanics of Motivated Reasoning*, 30 J. Econ. Persps. 133 (2016).
	Herding and Conformity	• Meng Liu et al., *Algorithm Aversion: Evidence from Ridesharing Drivers*, Mgmt. Sci. (2023) • Cass R. Sunstein, Conformity (2019). • Cass R. Sunstein, *The Use of Algorithms in Society*, Rev. of Austrian Econ., May 4, 2023. • Matthew Salganik et al., *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market*, 311 Science 854 (2006)

Category	Driver of Aversion	References in Existing Academic Literature
General Comfort and Skepticism	**Low Levels of Comfort with Innovation and Technology**	• Hasan Mahmud et al., *What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion*, 175 Tech. Forecasting & Soc. Change, Feb. 2022, at 1. • Maximilian Germann & Christoph Merkle, *Algorithm Aversion in Delegated Investing*, 93 J. Bus. Econ. 1691 (2023).
	Lack of Understanding and Trust	• Hasan Mahmud et al., *What influences* Hasan Mahmud et al., *What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion*, 175 Tech. Forecasting & Soc. Change, Feb. 2022, at 1. • Hasan Mahmud et al., *Decoding Algorithm Appreciation: Unveiling the Impact of Familiarity with Algorithms, Tasks, and Algorithm Performance*, 179 Decision Support Sys., Apr. 2024, at 1. • Michael Yeomans et al., *Making Sense of Recommendations*, 32 J. Behav. Decision Making 403 (2019). • Cass R. Sunstein, *The Use of Algorithms in Society*, Rev. of Austrian Econ., May 4, 2023 • Lingwei Cheng & Alexandra Chouldechova, *Overcoming Algorithm Aversion: A Comparison between Process and Outcome Control*, 2023 CHI Conf. on Hum. Factors Comp. Sys. 12982.
Task-Specific Confidence and Suitability	**Perceived Suitability for Algorithmic Decision-Making**	• Noah Castelo et al., *Task-Dependent Algorithm Aversion*, 56 J. Marketing Rsch. 809 (2019) • Hasan Mahmud et al., *What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion*, 175 Tech. Forecasting & Soc. Change, Feb. 2022, at 1. • Yoyo Hou & Malte Yung, *Who Is the Expert? Reconciling Algorithm Aversion and Algorithm Appreciation in AI-Supported Decision Making*, 5 Proc. ACM Hum.-Comp. Interaction 1 (2021) • Samantha Joel et al., *Is Romantic Desire Predictable? Machine Learning Applied to Initial Romantic Attraction*, 28 Psych. Sci. 1478 (2017) • Michael Yeomans et al., *Making Sense of Recommendations*, 32 J. Behav. Decision Making 403 (2019).

Category	Driver of Aversion	References in Existing Academic Literature
	High Levels of Confidence in Human Decision-Maker	• Cass R. Sunstein, *The Use of Algorithms in Society*, REV. OF AUSTRIAN ECON., May 4, 2023. • Victoria Angelova et al., *Algorithmic Recommendations and Human Discretion* (Nat'l Bureau of Econ. Rsch., Working Paper No. 31747, 2023) • Yoyo Hou & Malte Yung, *Who Is the Expert? Reconciling Algorithm Aversion and Algorithm Appreciation in AI-Supported Decision Making*, 5 PROC. ACM HUM.-COMP. INTERACTION 1 (2021) • Esther Kaufmann et al., *Task-Specific Algorithm Advice Acceptance: A Review and Directions for Future Research*, 7 DATA & INFO. MGMT., Sept. 2023, at 1. • Ryan T. Allen et al., *Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion*, 33 ORG. SCI. 149 (2022). • Cass R. Sunstein & Lucia Reisch, *On Liking Algorithms*, ENV'T & RES. ECON. (forthcoming 2024).
	Ignored Factors and Unintended Consequences	• Ryan T. Allen et al., *Algorithm-Augmented Work and Domain Experience: The Countervailing Forces of Ability and Aversion*, 33 ORG. SCI. 149 (2022).

III. INTERNAL FACTORS

A. Agency and Control

In some cases, people choose not to follow algorithmic advice because of a desire to maintain and exercise their own agency.[37] There are a variety of plausible drivers of this desire. People may believe that choice has intrinsic value, and they may want to choose for this reason.[38] If so, they might want to choose even if they might not choose well. Alternatively, they might be adopting a heuristic, the "I Know Best What Is Best For Me" Heuristic, and they might follow that heuristic even if it leads to serious errors. Or they might want to maintain responsibility for their choices. They might like the idea that if things turn out well, it is because they made the right choice; they might even like the idea that if things do not turn out well, it is because they made the wrong choice. In any case, the choice was theirs. They might also want to choose in order to learn. They might think that choosing is a muscle, and they might want to exercise it. They might know that they will err, but they might be willing to accept the cost of error if the result is to gain knowledge

[37] Sunstein, *supra* note 1, at 11; *See*, *e.g.*, Roy Shoval et al., *Choosing to Choose or Not*, 17 JUDGMENT & DECISION MAKING 768, 768-70 (2022); Sebastian Bobadilla-Suarez et al., *The Intrinsic Value of Choice: The Propensity to Under-Delegate in the Face of Potential Gains and Losses*, 54 J. RISK & UNCERTAINTY 187, 188 (2017); Mahmud et al., *supra* note 13, at 11-12.

[38] Bobadilla-Suarez, *supra* note 37, at 199 ("people are willing to pay a control premium to make their own choices").

about how to do better in the future. In practice, it might be difficult to distinguish between a "pure" desire to exercise agency, because of the intrinsic value of choice, and a desire to assume responsibility or to learn.

The general point is that in some cases, people choose to make their own judgment or follow their own intuition toward what might be a suboptimal outcome, rather than to follow an algorithm, because the act of choosing fulfills a desire for sovereignty over one's own life and generates utility or welfare for that person independent of the outcome of their decision. They prefer to be subjects rather than objects.

For example, someone who is planning a vacation might not want to rely on an algorithm, even if the algorithm will choose better than she will. Part of the fun of the vacation might involve planning for it. If someone is making a medical decision, she might want to make it herself, rather than rely on an algorithm, fearing a lack of control over her own body. Similarly, a retiree may pick investment options even if he realizes that doing so will likely yield lower returns as compared to following an algorithm or a robo-adviser.[39] A desire for agency, although admittedly unlikely to be widespread in this context, may drive a decision to use old-fashioned self-navigation instead of following a GPS or an app (Google Maps, Apple Maps, Waze, and so forth). Some people enjoy finding their own way even if it means spending a few extra minutes in traffic.

Although the phenomenon has not attracted much research interest, we speculate that an opposing desire may create algorithm appreciation. Some people prefer not to choose. They simply do not enjoy making (certain) decisions, finding it unpleasant or stressful to exercise agency.[40] They might want to avoid responsibility, not to assume it. For such people, the desire to avoid making (certain) decisions can lead to algorithm appreciation. A retiree who finds managing his finances unpleasant or overwhelming, or who does not feel comfortable deciding whether to follow or ignore the advice of human advisers, may be relieved to invest in ETFs or use a robo-adviser. People who do not like making their own medical decisions might greatly prefer to rely on an algorithm. People who do not enjoy navigation, or who find the task demanding or stressful, may use a GPS or an app to avoid exercising agency over their routes. The general empirical literature on when people choose to choose, or instead choose not to choose, helps explain when we will find algorithm aversion and when we will find algorithm appreciation.[41]

Note that in emphasizing agency, we are assuming that people are deciding between making their own decisions and relying on an algorithm, not between

[39] Salman Farooqui, *Despite tough times, it's been a good year for those who use robo-advisers*, THE GLOBE AND MAIL (Nov. 10, 2023), https://www.theglobeandmail.com/investing/personal-finance/household-finances/article-despite-tough-times-its-been-a-good-year-for-those-who-use-robo/ [https://perma.cc/QKZ6-8QGV].

[40] *See* Cass R. Sunstein, *Choosing Not To Choose*, 64 DUKE L.J. 1, 9 (2014) (noting that technical, complex choices are often undesirable).

[41] *See id.*

relying on another human being and relying on an algorithm. Note also that if people show algorithm aversion because they want to exercise their own agency, we can imagine possible corrective measures, if they are desired. For example, it might be helpful to emphasize that reliance on an algorithm will lead to substantially better outcomes, which might make it less appealing to exercise one's own agency. Those who seek to encourage the use of algorithms in the private or public sector might directly address the desire for agency in this way. Would people want to exercise agency if the consequence is to lose substantial sums of money, or to endanger their health? Answering this question would require balancing the benefit of exercising agency against the deterioration in outcomes.

B. Moral or Emotional Qualms

There are some decisions that people are uncomfortable delegating to algorithms, and not because they want to exercise their own agency. Even if an algorithm would perform better than a human decision-maker would, people may be reluctant to leave the final choice to an algorithm. One reason might be moral: there might be a moral judgment, or a moral intuition, that certain decisions ought to be made by human beings.[42] A decision whether to give someone bail or asylum, or what kind of criminal sentence to give to a defendant, might fall in this category. If a moral judgment of this kind is at work, it may or may not be thought through;[43] if it is, it might be a product of one or more of the mechanisms discussed below (such as a belief that an algorithm is likely not to consider some relevant variable). Another motivation might be more primitive: there might be a simple emotional sense that certain decisions ought to be made by a person.[44] Of course, an emotional reaction might be the source of a moral judgment or moral intuition, and it might have moral sources of some kind; it might even be a heuristic.[45] Neuroscientific research might be able to sort out some of the complexities here.

Algorithm aversion may arise or be heightened when outcomes are highly personal or in some sense sensitive, or when the decision is or seems grave. Some people may be uncomfortable with allowing an algorithm to decide whether someone from another country should receive a visa, or whether a defendant should be forced to stay in jail pending trial.[46] So, too, they might not like the idea that an algorithm will choose a treatment plan for a critically ill patient.[47] Although an algorithm may outperform human judges in projecting a defendant's flight risk[48] or

[42] *See, e.g.*, Kleinberg et al., *supra* note 28, at 241; Mullainathan & Obermeyer, *supra* note 29; Filiz et al., *supra* note 1.

[43] It might in fact be a moral heuristic. *See* Cass R. Sunstein, *Moral Heuristics*, 28 BEHAV. & BRAIN SCI. 531, 531-32 (2005).

[44] It is useful to think here about System 1—which operates in the mind automatically and quickly, with little or no effort and no sense of voluntary control—as a driver of judgments. *See* DANIEL KAHNEMAN, THINKING, FAST AND SLOW 21-32 (2011).

[45] *See* Sunstein, *supra* note 43.

[46] *See* Kleinberg et al., *supra* note 28.

[47] *See* Sunstein, *supra* note 1, at 11.

[48] *Id.* at 4.

be able to create medical treatment plans that better address the common needs of patients,[49] some people may be fundamentally uncomfortable with the idea that an algorithm, lacking empathy or emotions, should determine whether a person is to be freed or incarcerated, or should make a decision that could have life-or-death implications. Here again, we might be dealing with a moral heuristic.

In some cases, a correlation between gravity and algorithm aversion will present a "tragedy of algorithm aversion."[50] If algorithm aversion is especially prominent in certain highly important and sensitive decisions, and if algorithms tend to outperform human decision-makers, algorithm aversion differentially leads to suboptimal outcomes in some of the most important contexts.[51] It may lead to more crime, more imprisonment, or both. It may lead to more deaths from heart diseases.[52] Decisions that implicate people's health, safety, or freedom may be especially susceptible to bias and emotional responses, as people struggle to make life-altering choices, or struggle to evaluate how public officials should make such choices. Although algorithms may be less biased and noisy, feelings of moral and emotional gravity may pull people away from algorithms when making important decisions.

These points have evident implications for policy and law. An evident response might be to emphasize the data: If algorithms really would produce significantly better outcomes, perhaps we can reduce algorithm aversion.[53] Here too, an emphasis on accuracy, and on what is lost by not using algorithms, might combat the relevant aversion. Consider this point in the context of screening of travelers: if using an algorithm increases accuracy and reduces discrimination, as opposed to reliance on human beings, would people insist on relying on human beings? In short, showing robust data on algorithm performance could help overcome the algorithm aversion that arises from moral qualms.

C. Unique Skills

Suppose that someone has special or unique skills. A cardiologist might have been treating heart disease for over twenty years; an investment adviser might have been advising clients for a decade. A person who has extensive experience might be reluctant to defer to an algorithm. Such a person might believe, reasonably even if wrongly, that she will do better than any algorithm will or can. Alternatively, such a person might be engaging in motivated reasoning; she might refuse to recognize the superiority of the algorithm simply because it is deeply unpleasant to do so.[54]

[49] *See* Mullainathan & Obermeyer, *supra* note 29.

[50] Filiz et al., *supra* note 1, at 15.

[51] *See* Mullainathan & Obermeyer, *supra* note 29, at 723.

[52] *See* Ludwig et al., *supra* note 11, at 14-15.

[53] For preliminary data to this effect, see Sunstein & Reisch, *supra* note 27, at 15-17.

[54] *See* Nicolas Epley & Thomas Gilovich, *The Mechanics of Motivated Reasoning*, 30 J. ECON. PERSPS. 133, 133-39 (2016).

In some cases, algorithm aversion has been found to rest on a mechanism of this general kind.[55] If a person believes that she is uniquely talented or knowledgeable in some way, she may be motivated not to defer to an algorithm, or not to recognize its superior performance, because doing so would imply (1) that her expertise is not unique after all or (2) that new technology can improve on the skills that she built over time. (Again, algorithm aversion might be rational or even right in some such circumstances.[56]) These concerns might arise in "identity-relevant" situations,[57] which present questions about a person's notion of self. If I recognize that algorithms are better for these decisions, what would that imply for me? What would happen to me, and to people like me, if tasks like this were generally or always delegated to algorithms?

A perception of unique skills has largely been found in highly-experienced, knowledge-based workers[58] (e.g., doctors, lawyers, and IT professionals). Some people who have spent years (or decades) building a base of knowledge appear to show algorithm aversion. They might think: An algorithm cannot possibly perform as well as I can. Or they might think: If I have worked for my whole life to develop expertise in a subject, how could an algorithm possibly know more than I do? Although existing research has not focused much on skilled workers in physical trades, rather than knowledge-based professions, we speculate that such workers might have the same thoughts that cause algorithm aversion. The same factors of identity relevance and self-perception might apply to any skilled worker. If somebody suggested that an algorithm could diagnose and autonomously solve a set of common household plumbing problems better than a human plumber could, why would a plumber react any differently from IT workers tasked with fixing technological problems?[59] When a perception of unique skills is at work, algorithm aversion might be especially stubborn; it remains to be seen whether data on the superior performance of algorithms (if there is such data in the relevant context) might help. But exploring ways to overcome algorithm aversion in skilled individuals may be particularly worthwhile, as their expertise may allow greater synergy with algorithms in reaching optimal results.

It is important to note that the existing research on algorithm aversion is heavily focused on the western world, especially the United States.[60] There is some evidence that concerns about uniqueness are more relevant to algorithm aversion in

[55] *See, e.g.*, Allen et al., *supra* note 30; Mahmud et al., *supra* note 13, at 12; Nicole Tsz Yeung Liu et al., *Is algorithm aversion WEIRD? A cross-country comparison of individual-differences and algorithm aversion*, J. RETAILING & CONSUMER SERVICES, May 2023, at 1, 1-8; Noah Castelo, *Perceived corruption reduces algorithm aversion,* 34 J. CONSUMER PSYCH. 326, 327 (2023).

[56] *See* Victoria Angelova et al., *Algorithmic Recommendations and Human Discretion* 1-5 (Nat'l Bureau of Econ. Rsch., Working Paper No. 31747, 2023).

[57] Castelo, *supra* note 55, at 327 ("Algorithm aversion (or preference for humans) is strongest when the task is identity-relevant and when evaluative criteria are ambiguous.").

[58] E.*g.*, Allen et al., *supra* note 30.

[59] *Cf. Id.*

[60] *See* Liu et al., *supra* note 55, at 326.

the United States, and potentially other countries with highly "individualistic" cultures, than in other parts of the world with different cultural values and norms.[61]

D. Human Failure, Algorithmic Failure

Some evidence suggests that people are more willing to forgive human error than algorithmic error, and that people tend to penalize algorithms more for the same mistakes.[62] People expect and accept that other human beings will occasionally err and are often willing to forgive those errors and return for advice or help in the future. An error or piece of bad advice from an algorithm is more likely to discourage people from using the algorithm. It seems safe to say that differential forgiveness of human beings varies in magnitude among people, and those who penalize humans and algorithms similarly are less likely to show algorithm aversion.

To offer a bit more detail: Doctors, lawyers, investment advisers, and direction-givers all make occasional mistakes. Despite these mistakes, people may well be willing to return to the source of bad advice or decisions, recognizing that making mistakes is part of being human and that even the foremost experts occasionally err. On the other hand, some people are less willing to re-use an algorithm that has erred or given the same bad advice or made the same bad decision.[63] It is worth noting that when people face repeated circumstances—for example, people may face a new medical issue a few months after dealing with a misdiagnosis—they cannot simply avoid dealing with the issue repeatedly. Forgiveness of human errors may arise out of some degree of necessity in some cases. Faced with these situations, people tend to forgive human error more easily than algorithmic error and are more likely to return to a faulty human source.[64]

E. Loss Aversion, Risk Aversion

There is some evidence that an individual's level of aversion to risk or loss can help drive algorithm aversion.[65] If the proposed or projected outcomes presented in an algorithm's output represent losses for the decision-maker relative to the current state, the decision-maker may be less likely to rely on an algorithm rather than on his own judgment. On the other hand, the same loss-averse person

[61] *Id.* at 328-331.

[62] *See, e.g.*, Berkeley J. Dietvorst et al., *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCH. 114, 124 (2015); Alvaro Chacon et al., *A longitudinal approach for understanding algorithm use*, J. BEHAV. DECISION MAKING, OCT. 2022 at 1, 1-5.

[63] Chacon et al., *supra* note 62, at 10.

[64] *Id.*

[65] E*.g.*, Mahmud et al., *supra* note 13, at 12; Inga Toma et al., *Impact of Loss and Gain Forecasting on the Behavior of Pricing Decision-making*, 6 INT'L J. DATA SCI. & ANALYSIS 12, 14-18 (2020).

will be more likely to rely on an algorithm's advice when the algorithm projects a positive outcome.[66]

Note that this bias against algorithmically-advised losses is not, strictly speaking, about decision-makers, or about whether people are preferred to algorithms. Instead, the bias penalizes algorithms for pessimistic projections rather than inaccuracy. At the same time, this kind of penalty might turn into algorithm aversion if an algorithm projects losses while a human being projects gains. For that reason, a person who is less open to accepting and dealing with projected losses will be less likely to rely on algorithms when they project negative outcomes. For example, a loss-averse bond trader may have an algorithm that projects the next month's returns for a variety of potential trades and suggests an optimal approach. The trader will be more likely to follow the algorithm's advice when it projects positive returns. When the algorithm's optimal approach is projected to yield negative returns, loss aversion may lead the trader instead to follow his own judgment or that of a more optimistic human adviser, hoping to find positive returns where none are initially presented.

Risk aversion may interact similarly with algorithm aversion. A person's appetite for risk may drive algorithm aversion no less than a person's attitude toward potential losses. If people are extremely risk-averse, they may over-rely on conservative algorithmic projections. For example, an accountant tasked with creating financial projections is more likely to rely on an algorithm's conservative projections if she is more risk averse. If the same accountant is less risk-averse, she may be more open to substituting her own analysis for the algorithm's or replacing the algorithm's projections with more aggressive ones. Similar thinking can apply to junior employees who present analyses to their bosses. A risk-averse employee may think that following an algorithm's output rather than going out on a limb with creative analysis creates less career risk. On the other hand, algorithm appreciation may arise in situations where an employee looks to delegate potential blame or responsibility to an algorithm.

F. Confirmation Bias and Status Quo Bias

Some research finds that humans show confirmation bias when deciding whether to follow an algorithm's advice, becoming more averse when the advice conflicts with a person's existing intuition or long-held belief.[67] The more inconsistent an algorithm's advice is with a person's existing habits and beliefs, the less likely that person is to follow the advice. Here again, confirmation bias may or may not produce algorithm aversion, depending on whether the algorithm is disconfirming.

Presented with an algorithm that offered location recommendations, ridesharing drivers were found to be less likely to follow the algorithm's

[66] Mahmud et al., *supra* note 13, at 10; Toma et al., *supra* note 65, at 14-18.

[67] *See, e.g.*, Liu et al., *supra* note 31, at 6-10.

recommendation when it did not align with their past experiences.[68] Even though the algorithm was designed to optimize the matching of driver availability (supply) with passengers needing rides (demand), drivers were reluctant to forego their existing routines and ideas about how to best pick up rides.[69] It is worth noting, however, that the algorithm in the study was designed to optimize system-wide utilization rather than individual driver income.[70] The algorithm's design weakens any conclusion about algorithm aversion, for individual drivers may have in fact been better off optimizing for themselves rather than for the system.

At the same time, the study provides some basis for believing that confirmation bias can drive algorithm aversion. First, the algorithm in the study would optimize aggregate driver income by optimizing system-wide driver utilization[71]—so even if a few drivers would earn more by diverting from the algorithm, following the algorithm is likely the optimal course of action for many, and perhaps most, drivers. Second, it is not clear how the algorithm was explained to drivers or why they may have diverted. Considering the study's design and the fact that the researchers surveyed drivers about their attitudes toward algorithms,[72] we have some basis for inferring that the drivers' decisions were at least partially driven by algorithm aversion. More generally, existing work on confirmation bias makes it plausible to think that people may be more skeptical of algorithms when their judgments or advice differ from existing beliefs.[73]

There may be some overlap between confirmation bias and dedication to a routine, or status quo bias.[74] An algorithm that makes a recommendation inconsistent with a person's existing beliefs may also recommend actions that would necessitate a change in a person's consistent routine. Straying from routine may similarly make people reluctant to defer to an algorithm.

G. Herding and Conformity

Conformity pressures may contribute to algorithm aversion. People may be less likely to use an algorithm when it advises doing something that would cause them to stick out from their peers. Indeed, the same study of ride-sharing drivers found that drivers were less likely to follow the algorithm's advice when it would lead a driver to act differently from his peers.[75] When many drivers concentrated themselves around a single area or event, drivers were relatively unlikely to follow the algorithm's advice to go somewhere else. This behavior may come from a desire

[68] *Id.*

[69] *Id.*

[70] *Id.* at 10.

[71] *Id.*

[72] Liu et al., *supra* note 31, at 6-10.

[73] *See* Epley & Gilovich, s*upra* note 54, at 136.

[74] *See* William Samuelson & Richard Zeckhauser, *Status Quo Bias in Decision Making*, 1 J. RISK & UNCERTAINTY 7, 39 (1988) (discussing the basis of status quo bias, including, among others, an individual's tendency to defer to past decisions to guide decision-making).

[75] Liu et al., *supra* note 31, at 2.

not to look wrong or silly to around peers (reputational risk to oneself) or instead a belief in the wisdom of the (human) crowd (perceived risk that the algorithm is wrong).[76] Either way, it appears likely that the actions of others around us will often influence whether we use algorithms. In this sense, algorithm appreciation and aversion are likely influenced by peer pressure and the observations of others in the same way that hit songs are.[77]

IV. General Comfort and Skepticism

A. Low Levels of Comfort with Innovation and Technology

Algorithm aversion may result from anti-novelty biases and a general aversion to technology. Those who fear technological change and view algorithms as a symptom of that change are likely to be averse to algorithm use.[78] This aversion may manifest itself as a fear or skepticism of technology and/or of change more generally.[79] Either way, a person's level of comfort with adopting novel activities, processes, and tools may relate to their level of comfort with algorithms as a general matter. People who are broadly uncomfortable with algorithms will be more averse to algorithm use in a variety of situations,[80] independent of any other, more specific drivers of algorithm aversion that relate particularly to the decision at hand.

On the other hand, people who frequently deal with innovation and/or new technologies will likely be more comfortable with algorithm use. One study found relatively high adoption of a delegated investing algorithm among a test population of university students.[81] Consistent with our general point here, the authors emphasized that the experiment focused on young, highly educated, technologically inclined subjects, who may not be reflective of the broader population.[82]

B. Lack of Understanding and Trust

Some people have a general mistrust of algorithms and a lack of confidence in them[83] and will decide against using an algorithm in a given situation simply because they are unfamiliar with it.[84] This unfamiliarity is associated with a lack of understanding and hence trust, independent of a general discomfort with technology or the algorithm's performance. Even if people are generally

[76] *See* Cass R. Sunstein, Conformity 19-20, 23 (2019).

[77] *See* Sunstein, *supra* note 1, at 19-21; Matthew Salganik et al., *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market*, 311 Science 854, 854-856 (2006).

[78] *See* Mahmud et al., *supra* note 13, at 7.

[79] *See id.* at 7, 12.

[80] *Id.* at 7.

[81] Maximilian Germann & Christoph Merkle, *Algorithm Aversion in Delegated Investing*, 93 J. Bus. Econ. 1691, 1708 (2023).

[82] *Id.* at 1721.

[83] Mahmud et al., *supra* note 13, at 7.

[84] *See* Mahmud et al., *supra* note 25, at 8.

comfortable with new technologies and broadly aware of the superior performance of an algorithm, they may still carry a residual mistrust of the algorithm.

Human advice feels tangible and traceable. People often have a sense of where it comes from, on what it is based, and why the advice-givers prescribe as they do. To some, algorithms represent sketchy black boxes that are difficult to understand and therefore to trust.[85] People might not know how and why they work. They might be reluctant to follow advice that seems arbitrarily generated and will be more willing to rely on advice that seems properly grounded in familiar processes, including human reason.[86] Algorithm aversion can be driven by this reluctance.[87] People may mistrust algorithms because they do not understand how or why they work, independent of how well they actually work.[88] In this sense, trust and understanding go hand-in-hand.

One study found that giving a person control over the algorithm's testing process—in the form of choosing the training algorithm—creates a similar aversion-mitigating effect as giving a person the power to adjust the algorithm's output post-hoc.[89] This finding can be taken to corroborate the relevance of understanding and trust, especially if we believe that exposure to the model's training process is a good proxy for understanding.

Consider an algorithm that can project which jokes a given person will find funny.[90] The algorithm can outperform human beings, including that person's friends and family.[91] The algorithm can do so because it has a great deal of data on which jokes many people find funny and can match a person's preferences to those of numerous others with similar senses of humor.[92] You may initially distrust this algorithmic joke recommender in favor of advice from people who know you well. Why should an algorithm that cannot "understand humor" know what you will find funny better than your friends can? Many people think that way. But once they are given a better idea of why the algorithm works, they become more likely to trust and use it.[93]

Popular coverage of powerful, scary algorithms likely contributes to the general mistrust of black box algorithms. From addictive social media algorithms[94] to too-

[85] *See* Yeomans et al., *supra* note 33, at 404.

[86] *See id.*

[87] *Id.*

[88] *Id.*; Sunstein, *supra* note 1, at 12.

[89] Lingwei Cheng & Alexandra Chouldechova, *Overcoming Algorithm Aversion: A Comparison between Process and Outcome Control*, 2023 CHI CONF. ON HUM. FACTORS COMP. SYS. 12982, 12982.

[90] Yeomans et al., *supra* note 33, at 404; Sunstein, *supra* note 1, at 12.

[91] Yeomans et al., *supra* note 33, at 404.

[92] *Id.*

[93] *Id.*

[94] Clothilde Goujard & Gian Volpicelli, *EU hits Meta with new probe over 'addictive' algorithms harming children*, POLITICO (May 16, 2024), https://www.politico.eu/article/meta-hit-with-new-eu-probe-over-addictive-algorithms-harming-children/ [https://perma.cc/VF3M-JBLG].

accurate algorithms that seem to spy on users,[95] the algorithmic "boogeyman" is not hard to find. The association of the term "algorithm" with secretive, all-powerful mechanisms that seem to spit out highly accurate recommendations based on impenetrable methodologies likely contributes to the general mistrust of algorithms that we do not understand. Here as well, an understanding of the mechanisms behind algorithm aversion offers some strong clues about how to combat it.

V. TASK-SPECIFIC CONFIDENCE AND PERCEIVED SUITABILITY

A. Perceived Suitability for Algorithmic Decision-Making

In some situations, people believe that the task at hand is not well suited to algorithms. The most commonly recognized aspect of people's perceptions of task suitability is objectivity.[96] Some people believe that the more subjective a task is—the more it requires considerations of "humanity" rather than logic or computation—the worse an algorithm will perform.[97]

One study found that people are more algorithm averse—strongly preferring human advice to algorithmic advice and mistrusting algorithms—for tasks like recommending romantic partners, writing news articles, and composing songs.[98] On the other hand, people trusted algorithms and preferred them to human beings for advice regarding driving directions, data analysis, and weather forecasts.[99] The study also asked participants to rate the objectiveness of each task.[100] Tasks seen as more subjective were more likely to yield algorithm aversion.[101]

To consider other examples, it is easy to imagine decisions that feel highly personal with no self-evidently correct answer ripe for optimization. Consider the question of whether to rely on an algorithm to help find you a date. You might think, not unreasonably, that no algorithm will have the relevant information, which

[95] Elana Klein, *The Latest Online Culture War Is Humans vs. Algorithms*, WIRED (APR. 29, 2024), https://www.wired.com/story/latest-online-culture-war-is-humans-vs-algorithms/ [https://perma.cc/9M6N-3YFV] .

[96] *See e.g.*, Noah Castelo et al., *Task-Dependent Algorithm Aversion*, 56 J. MARKETING RSCH. 809, 809-825 (2019) (showing that people trust algorithms less for tasks that seem more subjective in nature); Mahmud et al., *supra* note 13, at 13; Yoyo Hou & Malte Yung, *Who Is the Expert? Reconciling Algorithm Aversion and Algorithm Appreciation in AI-Supported Decision Making*, 5 PROC. ACM HUM.-COMP. INTERACTION 1, 21 (2021) (reconciling past studies and noting that effect of task type is observable where the difference in perceived expertise between the human and algorithm is small).

[97] Castelo, *supra* note 96, at 821.

[98] *Id.* at 816

[99] *Id.* at 814-816.

[100] *Id.* at 814.

[101] *Id.* at 816.

depends on a range of highly personal factors.[102] Or consider the question of whether to ask an algorithm to name your child. Reliance on an algorithm might make sense if you have some specific goal in mind ("a name that no one else in my community/state/nation is likely to have"), but if you want a name that "feels right" to you, given the wide range of factors that matter to you, an algorithm might at best be an adviser. Naming a child is not an optimization problem like naming a stadium,[103] and it might require consideration of a host of factors (probably, and we hope, not including profit maximization). If you want to send your child (once you've chosen a name) to a sleepaway camp during the summer, you might want to consider the options and make a decision that feels unique to your child's needs and your desired environment for them. Rightly or wrongly, such decisions might feel highly subjective, value-laden, and person-specific.

But are some decisions really too subjective for algorithms? To answer that question, we need to specify the meaning of the word "subjective."[104] Perhaps it refers to a set of considerations that are unique to the preferences and values of the chooser, such that a population-wide average, or even a more narrowly described average (say, an average for the chooser's demographic), would be too crude or coarse to capture what the chooser most cares about. If there is a set of decisions that is so subjective, in that sense, that algorithms cannot properly identify the relevant considerations, are people good at assessing which decisions fall into the set?

It is worth noting that in the above-mentioned study that found a link between task objectivity and algorithm aversion (published in 2019) "predicting joke funniness," "recommending a romantic partner," and "recommending a gift" were among the tasks that participants believed were most subjective and least likely to inspire trust in algorithms.[105] It is an open question whether these perceptions would be different today from what they were in 2019. In 2024, we know that given a proper base of data inputs, algorithms can perform better than humans in predicting how funny a given person will find a joke.[106] A variety of dating services exist among the most popular smartphone apps, many of which operate on algorithms that recommend romantic partners based on previously demonstrated preferences

[102] *See* Joel et al., *supra* note 34, at 1486 (discussing the inability of an algorithm to predict relationship desire and noting the difficulty of identifying input measures that would predict relationship desire).

[103] *See* MICHAEL WEAVER, DUKE & PHELPS, ARE FOOTBALL STADIUM NAMING RIGHTS UNDERVALUED? A COMPARISON BETWEEN THE UK AND U.S. 2-6 (2019), https://www.kroll.com/-/media/assets/pdfs/publications/valuation/duff-and-phelps-stadium-naming-rights-2019.pdf [https://perma.cc/9ZX4-7KW7] (showing how naming a stadium is a complex problem that warrants consideration of a wide range of factors).

[104] *See Subjective*, MERRIAM-WEBSTER ONLINE DICTIONARY, https://www.merriam-webster.com/dictionary/subjective [https://perma.cc/TQM9-UBGV] (last visited Sep. 7, 2024) (definitions include "peculiar to a particular individual" and "modified or affected by personal views, experience, or background").

[105] Castelo, *supra* note 96, at 816.

[106] *Cf.* Yeomans et al., *supra* note 33, at 404.

and the preferences of similar users.[107] So far as we are aware, there is no clear evidence about the accuracy of such algorithms.

B. High Levels of Confidence in Human Decision-Maker

In some situations, people have a high level of confidence in a specific human decision-maker, thinking that they will perform better than an algorithm. In some cases, this belief may be accurate;[108] perhaps the relevant decision-makers outperform algorithms. As a general matter, and quite reasonably, people are more averse to deferring to an algorithm when the human alternative is perceived to be a good predictor, forecaster, or adviser with regard to the decision at hand.[109] When a person is described as an "expert" or believed to be highly skilled and/or capable at the task, we are more likely to find algorithm aversion.[110] On the other hand, when a person is described as a randomly chosen decision-maker or believed to be relatively unskilled or incapable, we are more likely to find algorithm appreciation.[111] People sometimes treat their own expertise similarly to the expertise of others;[112] if a person believes that she is a highly capable expert, she may be more likely to believe that she will make better decisions than an algorithm.

Some people are unlikely to choose an algorithm over a well-educated, licensed doctor,[113] but they are more likely to take an algorithm's diagnostic advice instead of that of a stranger they encounter on the street. People may be reluctant to accept an algorithm's recommendation of which player the New England Patriots should start at quarterback over that of the team's coach, but they may trust the algorithm's pick more than that of the stranger complaining about the team's performance on the Monday morning commuter train.[114]

Suppose we believe that as a general rule, algorithms can outperform most human decision-makers, but also that the best human decision-makers can outperform algorithms.[115] Could most people identify the best human decision-makers? Given the extensive educational and licensing requirements, it might seem relatively easy to identify a qualified doctor, and we can likely feel relatively confident in their expertise. Still, it may not be so easy to know which of the most qualified doctors outperform a relevant algorithm. Research has found that while

[107] *See The Technology Behind Popular Dating Applications*, CAPITOL UNIVERSITY: CAPITOLOGY BLOG (FEB. 14, 2024), https://www.captechu.edu/blog/technology-behind-popular-dating-applications [https://perma.cc/C3QH-4V6F].

[108] *See* Sunstein, *supra* note 1, at 9-10 (noting that in the context of bail decisions, 10% of judges outperform algorithms).

[109] *See* Sunstein & Reisch, *supra* note 27, at 6-15; Sunstein, *supra* note 1, at 13; Hou & Yung, *supra* note 96, at 20-22; Kaufmann et al., *supra* note 26, at 8.

[110] *See* sources cited *supra* note 109.

[111] *See* sources cited *supra* note 109.

[112] *See e.g.*, Allen et al., *supra* note 30, at 163-64.

[113] *See* Sunstein & Reisch, *supra* note 27, at 15.

[114] *Cf. id.*

[115] *See* Sunstein, *supra* note 1, at 9-10; Angelova et al., *supra* note 56, at 1-5.

algorithms outperform 90 percent of human judges in the context of bail decisions, the top 10 percent of judges outperform algorithms.[116] Is it possible and feasible to identify the top 10 percent of judges in real time to determine whether they should follow an algorithm's advice when making decisions on a defendant's bail conditions?

C. Neglected Factors and Unintended Consequences

Reasonably enough, people may be less inclined to use algorithms when they believe that they do not consider relevant factors, or that the algorithm's output misses a fundamental part of the picture. This issue may be considered a subset of the other task-specific drivers mentioned above or may be thought of as distinct from them. In some cases, people choose to ignore or overrule an algorithm because they believe that following its advice will lead to unintended consequences that the algorithm does not and cannot consider.[117]

One study found that corporate IT support professionals exhibited algorithm aversion in cases where they foresaw broader negative consequences to the organization's IT systems as a result of the algorithm's recommended course of action.[118] Longer-serving employees believed that their experience with the company allowed them to consider additional relevant factors and be aware of the potential knock-on effects of various actions.

As a hypothetical example, consider an algorithm that a company uses to recommend employees for promotion. Suppose that the algorithm takes into account a variety of data points on the performance of entry-level employees (including analytical, behavioral, and other factors) and offers a recommendation of two employees each year for promotion. Although the algorithm may accurately reflect performance and leadership capabilities, it may not (let us suppose) consider the fact that some employees will not fit in well with the existing management team or with other team members, some of whom may leave the company depending on its promotion decisions. Occasionally, the executive in charge of promotions might overrule the algorithm to avoid promoting somebody who would be a poor fit with the other managers and create undesirable harm to the business that the algorithm failed to consider.

VI. TWO OPEN QUESTIONS

Existing research leaves many questions open, with implications for law and policy and for how we define and assess algorithm aversion. Consider two.

[116] Angelova et al., *supra* note 56, at 1-5.

[117] *See* Allen et al., *supra* note 30, at 163-164.

[118] *Id.*

A. The Net Impact of Interpretability

There seems to be a trade-off as algorithms become more interpretable and people become more comfortable with them. As we have seen, people are more likely to use an algorithm when they understand how and why it works. At the same time, a person with a greater understanding of and comfort with an algorithm may be more likely to second-guess the algorithm and overrule it.[119] While greater interpretability drives algorithm appreciation, people who think that they understand an algorithm well enough to trust it may also feel that they understand the algorithm well enough to know when to ignore it. The problem is that greater interpretability does not necessarily lead to accuracy in users' assessments.[120]

If we assume that an algorithm will outperform a human decision-maker in a given task, increasing the proportion of instances that leverage the algorithm should increase the aggregate accuracy of all decisions.

Illustratively:

Table 3: Effect of Algorithm Interpretability on Aggregate Decision Accuracy

aggregate accuracy = **[*% of decisions made by algorithm × % algorithm accuracy*] +** **[*% of decisions made by humans × % human accuracy*]**	
In period 1	Algorithm is less interpretable
In period 2	Algorithm is more interpretable, causing: ● % of decisions made by algorithm to increase (including those overruled) ● % algorithm accuracy to decrease (because some are wrongly overruled)

While interpretability increases the propensity to use the algorithm, it may simultaneously decrease the average accuracy of algorithmically-informed decisions (including those where a person "uses" the algorithm but overrules its output). Mathematically, if we consider overruling an "algo decision" rather than a "human decision," interpretability would increase the proportion of decisions made via the more accurate method but decrease the difference in accuracy between methods. On the other hand, if we consider overruling a human decision, interpretability would cause some people to shift toward the more accurate method while others, empowered to overrule the algorithm, shift away from it. Either way, the net effect on accuracy is unclear and there is an open question about the net benefit of interpretability.

As a practical example, consider driving directions from a GPS app. Early on in the lives of these apps, familiarity with their algorithms was likely relatively low. Today, most users likely have some sense of how these apps use live traffic data to

[119] DeStefano et al., *supra* note 24, at 26-28.

[120] *See* Julia Cecil et al., *Explainability Does Not Mitigate the Negative Impact of Incorrect AI Advice in a Personnel Selection Task*, 14 Sci. Reps 1, 6-12 (2024).

project the time of potential routes and provide a recommended route based on its projections. Because users today find the algorithm more interpretable, they may rely on their understanding of the algorithm to overrule it—if, for example, they believe that their experience or intuition will lead to a faster route than the traffic data on which the app relies. The users who overrule the app's recommendation may be right, in the sense that their chosen route is faster. Alternatively, they may be wrong, increasing their travel time. The net impact on outcomes is unclear—and not extensively covered in existing research.

B. Hypothetical Situations and Real Life

Research suggests that people may well act differently in real-world situations compared to hypothetical studies, showing less aversion to using algorithms when their decisions have real-life consequences.[121] Suppose, for example, that an official is deciding whether to use algorithms in screening travelers for security risks. It might be that in a survey situation, people would prefer to rely on human judgment. But if the algorithm is accurate, would an official actually decline to rely on it? Emerging evidence suggests that when making actual judgments across a variety of scenarios, rather than responding to posed hypotheticals, people more consistently use algorithmic advice.[122] How may this finding inform the way we think about algorithm aversion? Does accurately measuring and analyzing algorithm aversion require more real-world field studies and fewer designed experiments that deal with hypotheticals? It would seem so. If there is a substantial difference in algorithm usage rates between hypothetical scenarios and actual decisions, does this phenomenon apply evenly across all situations or primarily in identifiable types of situations? The answer remains unclear—and it is one we ought to try to find.

There is a call to action here in the domain of research, but we have also signaled potential action in circumstances in which algorithm aversion is real. As noted, something like algorithm aversion is likely to be found in many circumstances in which people might be asked to rely on artificial intelligence (and resist doing so), and hence an understanding of algorithm aversion has broad applicability to emerging issues in law and policy. We should make a simple distinction here.

First: In some circumstances, algorithm aversion (or AI aversion) is rational and even appropriate. For example, algorithms might lack local knowledge, and human beings, or some of them, might outperform them. Or people might reasonably care about something other than or in addition to accuracy (say, taking responsibility for their own choices or lives), and if so, they might not want to rely

[121] *See* Jennifer Logg & Rachel Schlund, *A Simple Explanation Reconciles "Algorithm Aversion" and "Algorithm Appreciation": Hypotheticals vs. Real Judgments* 21-28, (unpublished working paper) (Feb. 2, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4687557 [https://perma.cc/79VL-64QQ].

[122] *Id.*

on algorithms (or AI) even if doing so will lead to more mistakes. We have emphasized such cases. In short, algorithm aversion need not always be a problem.

Second: In some circumstances, algorithm aversion is based on a lack of information or on some kind of bias. These are the cases in which algorithm aversion can produce serious harm and therefore warrants the greatest concern. In some situations, it might lead to illnesses or deaths. In others, it might lead to other kinds of damaging mistakes, as in the domains of criminal justice, tax audits, environmental policy, road safety, and immigration. We have suggested that an understanding of the specific source of (damaging) algorithm aversion can point the way toward corrective measures. Most broadly, clear and vivid demonstrations of the advantages of using algorithms might help to overcome biases or heuristics that mislead people to prefer human judgment. Making use of algorithms simple or in some sense the default could also be beneficial. Increasing ease of access and use can dramatically increase adoption.[123]

VII. CONCLUSION

People show algorithm aversion when they prefer human forecasters or decision-makers to algorithms even though algorithms generally outperform people in the relevant context. Algorithm aversion has other forms, as when people prefer human forecasters or decision-makers to algorithms in the abstract, without having clear evidence about comparative performance in the relevant context. Algorithm aversion can exist as well when people prefer human forecasters or decision-makers in circumstances in which people are demonstrably superior to algorithms. In such cases, algorithm aversion is of course unobjectionable.

Regardless of what form it takes, algorithm aversion has important implications for policy and law. For example, it can significantly affect the criminal justice system, medical care, the tax system, immigration, international travel, and national security more broadly.[124] Both public and private institutions are likely to be affected by algorithm aversion and perhaps to seek ways to reduce or eliminate it.

We have seen that algorithm aversion is a product of diverse mechanisms, most prominently including (1) a desire for agency; (2) moral or emotional qualms about judgment by algorithms; (3) a belief that certain human experts have unique knowledge, unlikely to be held or used by algorithms; (4) ignorance about why algorithms perform well; and (5) a more negative reaction to algorithmic error than to human error. An understanding of the various mechanisms provides significant clues about the boundary conditions of algorithm aversion. It also provides

[123] *See* Peter Bergman et al., *Simplification and Defaults Affect Adoption and Impact of Technology, but Decision Makers Do Not Realize This*, 158 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 66, 68 (2020).

[124] *See Say hello to the new face of security, safety and efficiency: Introducing Biometric Facial Comparison Technology*, U.S. CUSTOMS & BORDER PROT. (Sep 3, 2024), https://www.cbp.gov/travel/biometrics [https://perma.cc/43KG-K6N4]; Collection of Biometric Data From Aliens Upon Entry to and Departure From the United States, 85 Fed. Reg. 74162 (proposed Nov. 19, 2020) (to be codified at 8 C.F.R. pts. 215, 235).

significant clues about how to overcome it. If, for example, people do not know why algorithms perform well, providing information on that question can reduce or eliminate algorithm aversion.[125] And if people wrongly believe that human experts have unique knowledge, educating them about the superiority of algorithmic judgments, if they are indeed superior, should help correct that belief.[126] Overcoming algorithm aversion in these situations may lead to far better outcomes for important decisions.[127]

[125] *See* Yeomans et al., *supra* note 33, at 412.

[126] For some evidence to this effect, see Sunstein & Reisch, *supra* note 27, at 18 (finding that a high percentage of people have a clear preference for either human beings or algorithms unaffected by brief information favoring one or another).

[127] *See* Ludwig et al., *supra* note 11, at 17-18.