Florida State University Law Review

Volume 48 | Issue 3 Article 8

Spring 2021

AI Advice

Sean Hannon Williams
University of Texas

AI ADVICE

SEAN HANNON WILLIAMS*

ABSTRACT

This Article merges one of our most ancient technologies for the promotion of welfare—advice—with some of our most recent—artificial intelligence (AI) and big data. AI is already writing novels, poetry, newspaper articles, and tweets. Big data may soon capture not only our online activities, but also our real-time heart rate, sleep patterns, and even our current mood. This is the first Article to introduce and examine the possibility of AI advice. AI advice offers the potential for exceedingly accurate personalized recommendations. It also reveals important limits within the burgeoning literature on personalized law. The Article first rejects recent attempts to rehabilitate mandatory disclosures by personalizing them. Ironically, the technological progress required to create effective big data disclosures will itself substantially reduce the need for such disclosures. In this future, advice, not disclosure, will be the dominant paradigm. The Article then dissects our everyday practices of advice-giving to unearth a number of powerful features of advice that promote self-efficacy, reduce motivated reasoning, and make it more likely that people will hear and heed good advice. The capacity to bundle these features with exceedingly accurate recommendations makes AI advice a promising alternative to its two main regulatory rivals: mandatory disclosure and nudges.

* F. Scott Baldwin Research Professor in Law, University of Texas School of Law. For insightful comments and conversations about this Article, I would like to thank Oren Bracha, Joseph Fishkin, William Forbath, Susan Morse, and Lior Jacob Strahilevitz.

INTRODUCTION

Advances in big data and artificial intelligence (AI)[1] point to a future where many regulatory efforts by the state can be personalized.[2] For example, scholars have suggested personalized speed limits sent to each person's smart car in real time, personalized consumer protections, and personalized disclosures.[3]

This Article introduces the concept of personalized AI advice. A sophisticated AI might be able to predict your personality traits, risk preferences, and goals by sorting through your internet search queries,

1. The term big data loosely describes the process of collecting and analyzing massive amounts of data to unearth previously unseen patterns. The term artificial intelligence refers to a collection of technologies that enable machines to learn, adapt, and solve problems in ways similar to humans. These technologies include neural networks, machine learning, cognitive computing, and natural language processing. 1 INFORMATION LAW § 1:13.

2. There has been an explosion of interest in personalizing law over the last decade. *See* Cass R. Sunstein, *Deciding by Default*, 162 U. PA. L. REV. 1, 10 (2013) (arguing that "personalized default rules are the wave of the future"); Ariel Porat & Lior Jacob Strahilevitz, *Personalizing Default Rules and Disclosure with Big Data*, 112 MICH. L. REV. 1417, 1418 (2014) (personalized default rules and disclosures); Omri Ben-Shahar & Ariel Porat, *Personalizing Negligence Law*, 91 NYU L. REV. 627, 629, 636-46 (2016) (negligence standards); Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards*, 92 IND. L.J. 1401, 1412-17 (2017) (speed limits and medical malpractice rules); Philipp Hacker, *Personalizing EU Private Law*, 3 EU. REV. PRIVACY L. 651, 669 (2017) (disclosures); Omri Ben-Shahar & Ariel Porat, *Personalizing Mandatory Rules in Contract Law*, 86 U. CHI. L. REV. 255, 255 (2019) (mandatory contract rules); Christoph Busch, *Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law*, 86 U. CHI. L. REV. 309, 309-12 (2019) (disclosures and privacy defaults); Adi Libson & Gideon Parchomovsky, *Toward the Personalization of Copyright Law*, 86 U. CHI. L. REV. 527, 528-29 (2019) (copyright penalties); Anthony J. Casey & Anthony Niblett, *A Framework for the New Personalization of Law*, 86 U. CHI. L. REV. 333, 347 (2019) (smart traffic lights); Matthew B. Kugler & Lior Jacob Strahilevitz, *Assessing the Empirical Upside of Personalized Criminal Procedure*, 86 U. CHI. L. REV. 489, 490 (2019) (Miranda warnings); *see also* Andrew Verstein, *Privatizing Personalized Law*, 86 U. CHI. L. REV. 551, 558 (2019) ("[T]rends in data gathering and analysis suggest that well-resourced lawmakers may soon have the technical ability to link directives to highly particular individual traits.").

3. *See infra* Section I.A.

social media networks, credit card purchasing history, and cell-phone use patterns. It may be able to predict your current mood by analyzing heart rate data from your smart watch, your facial expression at the ATM, and this morning's data from your internet-connected coffee pot.[4] An optimistic view of technological progress, combined with a pessimistic view of privacy protections, generate the conditions where an AI could someday predict whether you would be happier after getting cosmetic surgery, how much you should save for retirement, whether you should get the extended service warranty on your new TV, and even which engagement ring to buy.[5] It would then send personalized advice to each and every person in real time.

To date, the potential benefits of advice have been hidden by the current limits of creating good advice. Simply put, advice is often bad. But as AI benefits from big data (and potentially also from running a dizzying number of experiments every day to hone its causal inferences), those limits will recede. Accurate AI advice will become a real possibility, perhaps sooner than we think.

AI advice has powerful advantages over another futuristic regulatory tool: big data disclosures. These are disclosures that are personalized based on machine learning techniques sifting through massive amounts of information about you and your situation. Big data disclosures have attracted a great deal of attention because they offer a potential response to recent and powerful critiques of mandatory disclosure.[6] As those critiques have made clear, mandatory

---

4. *See infra* Section I.B.

5. Decision aids rooted in AI are already being explored in various professional contexts. Doctors, for example, might use AI to help identify health concerns in diagnostic images. Apoorva Mandavilli, *These Algorithms Could Bring an End to the World's Deadliest Killer*, N.Y. TIMES (Nov. 20, 2020) https://www.nytimes.com/2020/11/20/health/tuberculosis-ai-apps.html?action=click&module=RelatedLinks&pgtype=Article [https://perma.cc/KTM4-3BDS] ("Among all of the applications of A.I., . . . digitally interpreting an image using an algorithm instead of a human radiologist is probably furthest along."); Michael Matheny et al., *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*, NAT'L ACAD. MED. 39-40, 65, 94 (2020) (discussing diagnostic images and AI analysis of contract terms to assist lawyers); *see also* Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135, 1156 (2019) (arguing that AI "would at least be useful as an advisor to a [human] judge, especially if--as suggested above--it can lay out a set of reasons for the factual results it reaches"). This Article seeks to democratize these benefits. Everyone, not just experts, may someday benefit from AI advice or other forms of augmented intelligence.

6. Hacker, *supra* note 2, at 670 ("[A]t a time when disclosure is increasingly criticized as ineffective, personalization presents a way to strip disclosures of their excessive length and complexity."); Busch, *supra* note 2, at 330 ("Maybe the reports about the death of disclosures are greatly exaggerated. . . . [A]rtificial intelligence and superhuman information processing capabilities could redefine the optimal complexity of legal rules and refine, for example, the content of disclosures to a hitherto unachievable level of granularity."). This Article could have used the term AI disclosures, but that would abandon the alluring alliteration of big data disclosures. *See generally* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147 (2017); Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. CHI. L. REV. 1311 (2019).

disclosure is a problematic paradigm.[7] For example, literacy and numeracy problems are staggering.[8] But big data disclosures offer the potential to tailor messages for each individual's literacy and numeracy.

Advice is better at overcoming these barriers.[9] Advice does not require that you understand interest rates or complex decision trees. It can be as simple as "don't buy the extended service warranty." This requires no numeracy and very little literacy to understand. It also largely relieves you of the burden of decision fatigue and makes any difficult tradeoffs for you.

Consider the following analogy: current mandatory disclosure regimes create a frustrating game of connect the dots. The government provides you with large amounts of data (the dots) and leaves it to you to decide, first, which dots are relevant to you, and second, how to connect those dots in order to create some picture of what the world would look like if you made choice X. Big data disclosures reduce the number of dots. They do so by predicting which information is most relevant to you. They can only do so if the AI has some sense of what the world would look like if you made choice X. The interest rate on a loan may be particularly relevant to you, and the late payment penalties might not. The AI knows this because of its analyses of big data, which revealed your personality traits, income, spending habits, etc. The AI is already painting pictures that show the effects of choosing a high-interest loan or the effects of choosing a loan with a large late payment penalty. But once the AI has painted these pictures for itself, it seems somewhat cruel to offer you only a connect-the-dots version. Why not just tell people that choice X is better? This is what advice does.

In addition to these welfare-oriented benefits, advice is more autonomy-respecting than disclosure. Both seek to improve, rather than supplant, consumer decision-making. But people crave advice, not disclosure. In the medical context, for example, numerous studies have asked real patients about their real medical preferences. These studies consistently find that 50 to 80 percent of patients *do not want to decide for themselves*.[10] They actively want their doctor to decide for

7. *See generally* Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647 (2011); OMRI BEN-SHAHAR & CARL E. SCHNEIDER, MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE (2014).

8. BEN-SHAHAR & SCHNEIDER, *supra* note 7, at 79-86. *See infra* Section II.A.

9. *See id.* at 13, 185 (suggesting in passing that consumers want advice not disclosure).

10. See Section II.A. for a discussion of these studies.

them. Advice, not disclosure, much more closely reflects these preferences.

AI advice also has powerful advantages over classic nudges like default rules.[11] One enduring critique of using default rules to nudge people in particular directions is that they are not transparent and border on manipulation.[12] Advice, almost by definition, is transparent about its ultimate goal (to influence behavior) and its method (rational persuasion). The advice paradigm therefore avoids many of these critiques.[13]

In order to illustrate the advantages of the advice paradigm, this Article dissects our normal everyday practices of advice-giving. Natural advice—the kind you give a friend, or the kind you seek from family members—has a number of noteworthy features. As mentioned above, advice circumvents literacy and numeracy issues, relieves decision fatigue, and can help people make difficult tradeoffs. But natural advice does much more.

Natural advice has never been just about bare recommendations. Advice is often accompanied by words and actions that promote the advisee's self-efficacy, enhance the advisee's self-control, and preemptively disarm knee-jerk psychological defensiveness. Advice also has features of an ongoing service rather than a one-time consumable product. Advice can be bundled with precommitment mechanisms, monitoring, and reminders. These are all part of the "packaging" of natural advice. AI advice can mimic much of this welfare-enhancing packaging.

The Article is organized as follows. Part I provides an overview of AI, big data, and the recent scholarship on personalized law. This Part also illuminates a core irony of big data disclosures. Part II dissects natural advice to reveal a set of advantages that the advice paradigm has over both nudges and disclosures. It compares these tools along two dimensions: the extent to which they enhance welfare and the extent to which they preserve autonomy.[14] Part III maps out how policy makers could incorporate the advantages of natural advice into AI advice. Part IV addresses two potential objections. It argues that AI advice will not make us dumber and will not undermine our autonomy. This Part also argues that, even if policymakers ultimately reject AI

11. *See* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 35 (2008) (discussing the power of default rules).

12. *See infra* Section II.F. for further discussion.

13. *Id.*

14. This Article shares the same goal as many works of behavioral law and economics, to find tools between mandates and traditional disclosures that produce an appealing balance of welfare and autonomy. Ryan Bubb & Richard H. Pildes, *How Behavioral Economics Trims Its Sails and Why*, 127 HARV. L. REV. 1593, 1604 (2014).

advice, dissecting natural advice yields important insights for the design and scope of disclosures, nudges, and other regulatory messages.

I. BIG DATA AND AI

The term big data loosely describes the process of collecting and analyzing massive amounts of data to unearth previously unseen patterns.[15] Data is plentiful today. Facebook and Google already collect massive amounts of information about us. The "internet of things" allows many everyday objects like our cars or coffee makers to transmit data about us and our routines to companies.[16] In 2012, there were 9 billion devices of this sort.[17] Within the next year there may be 50 billion.[18] These include smart watches, step counters, thermostats, lights, refrigerators, ovens, glucose monitors, blood oxygenation sensors, baby monitors, security cameras, doorbells, door locks, helmets, soccer balls, cars, and much more.[19] It is estimated that these devices generate 500 trillion gigabytes of data each year.[20]

Advances in AI and machine learning techniques help sort through this unfathomably-immense and ever-growing trove of data.[21] These

15. Porat & Strahilevitz, *supra* note 2, at 1434-35.

16. Melissa W. Bailey, *Seduction by Technology: Why Consumers Opt Out of Privacy by Buying into the Internet of Things*, 94 TEX. L. REV. 1023, 1027-31 (2016).

17. Arbia Riahi Sfar et al., *A Roadmap for Security Challenges in the Internet of Things*, 4 DIGITAL COMM. & NETWORKS 118, 118 (2018).

18. Ted Saarikko, Ulrika H. Westergren & Tomas Blomquist, *The Internet of Things: Are You Ready for What's Coming?* 60 BUS. HORIZONS 667, 667-68 (2017).

19. Bailey, *supra* note 16, at 1027-31; Casey Crane, *The Internet of Things: 20 Surprising IoT Statistics You Don't Already Know*, THE SSL STORE (Sept. 4, 2019), https://www.thesslstore.com/blog/20-surprising-iot-statistics-you-dont-already-know/ [https://perma.cc/7UBR-8RUC]; Marie C. Baca, *Baby's First Smart Diaper: Pampers Takes 'Wearables' to a Whole New Level*, THE WASH. POST (July 18, 2019, 9:34 PM), https://www.washingtonpost.com/technology/2019/07/19/babys-first-smart-diaper-pampers-takes-wearables-whole-new-level/ [https://perma.cc/S5MF-5P4G]; Kai-Stefan Schober, *Start-Up from Dubai About to Increase Safety and Improve Workflows in the Building Industry*, ROLAND BERGER (Jan. 19, 2019), https://www.rolandberger.com/en/Point-of-View/Introducing-helmets-with-IoT-technologies-for-construction-sites.html [https://perma.cc/M3RV-VU8V]; Ketan Raval, *How IoT Industry Will Revolutionize Sports Equipment Industry?*, LET'S NURTURE (May 10, 2019), https://www.letsnurture.com/blog/how-iot-industry-will-revolutionize-sports-equipment-industry.html [https://perma.cc/ERH4-R5SX].

20. Melissa Liton, *How Much Data Comes from the IoT?*, SUMO LOGIC (Feb. 7, 2018), https://www.sumologic.com/blog/iot-data-volume/#:~:text=IoT%20data%20is%20measured%20in,to%20grow%20exponentially%2C%20not%20linearly [https://perma.cc/7G7P-UZPW].

21. Bernard Marr, *How Much Data do we Create Every Day? The Mind-Blowing Stats Everyone Should Read*, FORBES (May 21, 2018, 12:42 AM), https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#32aa704260ba [https://perma.cc/X52T-AR7F] (noting that we produce 2.5 quintillion bytes of data per day).

technologies—which I will refer to as "AI" as a shorthand[22]—have already proven themselves to be quite capable. AI can beat the best human players at chess, Go, and Jeopardy.[23] Of course, those games have far fewer variables than the real world. But AI has proven its potential in many other areas as well. A staggering 80 to 90 percent of trades on the US stock market are initiated by AI, not people.[24] AI can detect skin cancers as well as dermatologists.[25] AI can predict divorce based on subtle inflections of voice, and can now do so as well as trained therapists who have access to far more data.[26] Television broadcasters have allowed AI to generate entire highlight reels for sporting events.[27] Newspapers use AI to generate simple stories and tweets.[28] Companies use AI to automatically respond to customer inquiries.[29] Apps now boast the ability to predict, based on text messages and emails, the sender's personality traits and the probability that they are romantically interested in you.[30] Other apps use AI to predict when and whether you will incur overdraft fees from your bank.[31] AI has written novels, poetry, and even a relationship

22. 1 INFORMATION LAW § 1:13 ("'Artificial intelligence' broadly conceived is a constellation of technologies based on complex software that enables machines to 'learn,' adjust to circumstances, and to perform human-like tasks. It includes machine learning, deep learning reliant on neural networks, cognitive computing, computer vision and perception, and natural language processing, among other technologies.").

23. Casey & Niblett, *supra* note 2, at 1424.

24. *See* Silvia Amaro, *Sell-Offs Could be Down to Machines That Control 80% of the US Stock Market, Fund Manager Says*, CNBC (Dec. 5, 2018, 7:33 AM), https://www.cnbc.com/2018/12/05/sell-offs-could-be-down-to-machines-that-control-80percent-of-us-stocks-fund-manager-says.html [https://perma.cc/5TYG-UJNL] (reporting that approximately 80% of trades in US stock markets are made by algorithms, not humans, and that only 10% of JP Morgan's trades are initiated by its human employees).

25. Andre Esteva et al., *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115, 115 (2017).

26. *See* M.D. Nasir et al., *Predicting Couple Therapy Outcomes Based on Speech Acoustic Features*, 19 PLOS ONE 1 (2017).

27. At Wimbledon, computers analyze video feeds of the players, identify what the high points of the match were, and automatically generate a highlight reel. Brenda Salinas, *Wimbledon, Steeped In Tradition, Embraces Artificial Intelligence*, NPR MORNING EDITION (July 9, 2019, 5:05 AM), https://www.npr.org/2019/07/09/739784008/wimbledon-steeped-in-tradition-embraces-artificial-intelligence [https://perma.cc/V2VU-L2QA].

28. Bernard Marr, *Artificial Intelligence Can Now Write Amazing Content—What Does That Mean for Humans?*, FORBES (Mar 29, 2019, 1:23 AM), https://www.forbes.com/sites/bernardmarr/2019/03/29/artificial-intelligence-can-now-write-amazing-content-what-does-that-mean-for-humans/#5fb9e12c50ab [https://perma.cc/UTE6-QTX3].

29. Thomas L. Friedman, Opinion, *A.I. Still Needs H.I. (Human Intelligence), for Now*, N.Y. TIMES (Feb. 26, 2019), https://www.nytimes.com/2019/02/26/opinion/artificial-intelligence.html [https://perma.cc/G9CP-ZE6E].

30. KEIGO, https://keigo.app/ [https://perma.cc/V69G-RBG9] (last visited May 12, 2021) (predicting personality traits); Arielle Pardes, *Flirty or Friendzone? New AI Scans Your Texts for True Love*, WIRED (Sept 16, 2019, 7:00 AM), https://www.wired.com/story/ai-apps-texting-flirting-romance/ [https://perma.cc/6VEX-FF4R] (predicting romantic interest).

31. Natasha Lomas, *Dave Wants to Save You from Expensive Overdraft Fees*, TECHCRUNCH (April 25, 2017, 7:38 AM), https://techcrunch.com/2017/04/25/dave-wants-to-save-you-from-expensive-overdraft-fees/ [https://perma.cc/6MLF-WDTH].

advice column.[32] The New York Times offers a useful quiz to see if readers can distinguish between AI and human-created content; it's not easy.[33]

Although AI has proven quite capable, it has a long way to go before it can significantly personalize disclosures or offer accurate advice in a host of complex domains. But like other articles before it, this Article assumes an optimistic stance on technological advancement. "People often overestimate what will happen in the next two years and underestimate what will happen in ten."[34] This is likely true of our estimations of technological progress generally and AI specifically. Moore's law originally predicted that the number of transistors that would fit on a computer circuit would double every two years.[35] At the time, he was skeptical that his prediction could hold, but it did.[36] Moore's law has been applied more metaphorically to illustrate the rapid pace of advancements in other areas of computing.[37] In a classic example of this type of growth, an inventor introduces a King to a new game: Chess.[38] The inventor asks the King to buy the game for one grain of rice on the first square, then to double that amount for each square.[39] The King agrees, not realizing that by the time he reaches the 64th square, he will owe 18,446,744,073,709,551,616 grains.[40]

32. *See* Steven Poole, *The Rise of Robot Authors: Is the Writing on the Wall for Human Novelists?*, THE GUARDIAN (March 25, 2019, 4:00 AM), https://www.theguardian.com/books/2019/mar/25/the-rise-of-robot-authors-is-the-writing-on-the-wall-for-human-novelists [https://perma.cc/D996-SSA4]; Kata Karáth, *AI Agony Aunt Learns to Dole Out Relationship Advice Online*, NEW SCIENTIST (Jan. 26, 2017), https://www.newscientist.com/article/2119347-ai-agony-aunt-learns-to-dole-out-relationship-advice-online/#ixzz604nngMm5 [https://perma.cc/JH55-BLEH].

33. *See generally Sunday Review | Quiz: Did a Human or a Computer Write This?*, N.Y. TIMES (Mar. 7, 2015), https://www.nytimes.com/interactive/2015/03/08/opinion/sunday/algorithm-human-quiz.html [https://perma.cc/UYF6-QSR6] (asking, for example, whether the following was written by an AI: "Tuesday was a great day for W. Roberts, as the junior pitcher threw a perfect game to carry Virginia to a 2-0 victory over George Washington at Davenport Field"; it was.). Of course, not all AI content is worth printing. *See* Beth Skwarecki, *How to Train Your Own Neural Network*, LIFEHACKER (June 7, 2018, 10:30 AM), https://lifehacker.com/we-trained-an-ai-to-generate-lifehacker-headlines-1826616918 [https://perma.cc/4REY-A2NG] (offering these hypothetical titles for articles, generated by a simple AI: "The Best Way to Make Your Own Ink Out of the Winter," "How to Keep Your Relationship With an Imaginary Concept," and "The Best Way to Make a Perfect Cup of Wine With a Raspberry Pi").

34. BILL GATES, THE ROAD AHEAD 316 (1996).

35. Christopher S. Yoo, *Moore's Law, Metcalfe's Law, and the Theory of Optimal Interoperability*, 14 COLO. TECH. L.J. 87, 90 (2015).

36. *Id.*

37. *Id.* at 91.

38. *Grains on the Chessboard*, 40 NATURE GENETICS 1261, 1261 (2008).

39. *Id.*

40. *Id.*

These examples of exponential growth suggest that AI could soon offer powerful new levels of analysis

In addition to taking an optimistic view of technological progress, this Article will take a pessimistic view of privacy protections.[41] That is, it will assume that big data capacities will continue to expand, gathering more and more information about each person. For those who lament this hypothetical state of the world, the productive potential of AI advice might offer some solace.

A. The Power of Personalization

The potential of AI and big data has led a number of people to explore the benefits and perils of this future. Anthony J. Casey and Anthony Niblett envision a world where the government creates personalized rules for different people.[42] For example, everyone's cell phone might display a personalized speed limit that applies only to them and only in the present circumstances.[43] When weather conditions change, or when the AI obtains data that indicate you are sleepy, it will alter your personalized speed limit accordingly.[44] Omri Ben-Shahar and Ariel Porat have extended this to mandatory contract rules.[45] For example, different consumers might get mandatory cooling off periods of different lengths.[46] Ariel Porat and Lior Strahilevitz have applied a similar logic to default rules.[47] For example, if AI combs through data and finds that most heterosexual vegan men with Ph.Ds. in philosophy take their wife's last name, then this could be the default legal regime that governs those people.[48]

AI and big data could also lead to personalized disclosure regimes. Porat and Strahilevitz briefly discuss the possibility of using big data to personalize mandatory disclosures in medical and consumer settings.[49] For example, if a particular patient has a low risk of a particular side effect, an automated disclosure would omit reference to it and instead highlight information particularly relevant to her.[50]

---

41. It does not address potential privacy issues arising from an AI sifting through massive amounts of data about you, or accidentally revealing data about you. *See, e.g.*, Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG. (Feb. 16, 2012), https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [https://perma.cc/UN9V-CMCK] (describing the now-famous example of Target sending pregnancy related ads to people before they revealed their pregnancy to others). Instead, this Article highlights the positive potential of what many might feel are intrusive technologies.

42. Casey & Niblett, *supra* note 2, at 1404.

43. *See id.*

44. *Id.*

45. Ben-Shahar & Porat, *supra* note 2, at 256-57.

46. *Id.*

47. Porat & Strahilevitz, *supra* note 2, at 1465.

48. *Id.* (their example).

49. *Id.* at 1444-45.

50. *Id.* at 1445.

Most recently, Christoph Busch expanded on the concept of personalized disclosures by offering more potential illustrations of their utility.[51] For example, a customer interested in travel insurance who traveled to Libya last year (as evidenced by credit card data) would be told to pay particular attention to clauses that affect terrorism-related losses or losses overseas.[52] Similarly, once an AI learns a consumer's retail spending and return habits, it might design warnings to help avoid buyer's remorse.[53]

Big data disclosures have the potential to rescue mandatory disclosure from a set of powerful critiques recently levied against it by Omri Ben-Shahar and Car Schneider.[54] For example, one such critique of mandatory disclosures is that consumers are so overloaded with them that they could not possibly read them all.[55] If disclosures were shorter and more relevant—which personalization might accomplish—they would take less time and energy to process, and people might be more likely to read them.[56] The complexity of the disclosure could even be altered to align better with the consumer's cognitive capacities and personality.[57]

B. The Irony of Big Data Disclosures

There is an irony surrounding reliance on big data disclosures: The technological progress required to create effective big data disclosures will itself substantially reduce the need for such disclosures. In this future, advice, not disclosure, will be the dominant paradigm.[58]

Disclosure is appropriate when, for example, the disclosing entity knows something about population averages, but knows little about each individual. A city might know that many people overestimate the probability that they will repay a loan. In response, the city might mandate certain disclosures in an attempt to correct this misperception. But the city would make no claim to know how much each potential consumer needed the money, or whether the intended use for the money was sufficiently welfare-enhancing to justify the

51. Busch, *supra* note 2, at 315-16.

52. *Id.* at 316.

53. *See id.* at 318.

54. For a summary of those critiques, *see* BEN-SHAHAR & SCHNEIDER, *supra* note 7, at 56-57.

55. *Id.* at 101-05.

56. Porat & Strahilevitz, *supra* note 2, at 1472; *but see* Busch, *supra* note 2, at 322 (questioning whether personalized disclosures can do much to mitigate overload and decision fatigue problems).

57. Hacker, *supra* note 2, at 669 (cognitive capacity); Porat & Strahilevitz, *supra* note 2, at 1471, 1473 (personality).

58. Some scholars have highlighted a different potential irony. If people realize that the government has access to vast amounts of data about them, they will become more sensitive to privacy concerns and stop sharing their data. Niva Elkin-Koren & Michal S. Gal, *The Chilling Effect of Governance-by-Data on Data Markets*, 86 U. CHI. L. REV. 403, 407 (2019).

risk. This extra information is in the hands of the consumer alone. Disclosure allows each set of information to come together to allow the consumer to make an informed choice. So, disclosure works best when the city is well informed about some facts—interest rates, product attributes, etc.—but ignorant of other facts—the consumer's unique borrowing constraints, her other options, her preferences, etc.

AI and big data erode the consumer's informational advantage. The future of AI and big data may well allow entities like corporations and governments to know us better than we know ourselves. More than a decade ago, Oren Bar-Gill argued that credit card companies and cell phone companies know us better than we know ourselves, at least along limited dimensions.[59] For example, they might more accurately predict our probability of incurring penalty fees than we could. Today, companies know far more. Our search histories reveal many of our plans and preferences. Our social media activity and cell-phone use patterns reveal enduring features of our personalities.[60] Combined, these data may reveal much about our preferences, goals, aspirations, and plans.[61] Other data might reveal our day-to-day moods. Fitbits and Apple Watches already track our heart rates, exercise patterns, and sleep schedules. Our Tesla may report that we are driving aggressively, and our smart refrigerator might have data that suggests we ate too much sugar for breakfast. One might even imagine AI sifting through video camera feeds, analyzing your body language in real time, and assessing your current mood.[62] These are just a few examples that suggest we may not hold our informational advantage for much longer.

These forces can, of course, make big data disclosures better. But how will an AI know which disclosures are most relevant? In part, it will have to predict our goals and preferences. If the AI knows that I

---

59. Oren Bar-Gill & Franco Ferrari, *Informing Consumers about Themselves*, 3 ERASMUS L. REV. 93, 95 (2010); *see* Oren Bar-Gill, *Seduction by Plastic*, 98 NW. U. L. REV. 1373, 1419, 1429 (2004); *see also* Gerhard Wagner & Horst Eidenmüller, *Down by Algorithms? Siphoning Rents, Exploiting Biases, and Shaping Preferences: Regulating the Dark Side of Personalized Transactions*, 86 U. CHI. L. REV. 581, 582 (2019) ("[B]usinesses know at least as much about consumers as consumers know about themselves, and sometimes even more.").

60. Gokul Chittaranjan et al., *Mining Large-Scale Smartphone Data for Personality Studies*, 17 PERS. & UBIQUITOUS COMPUTING 433, 449 (2013); Yoram Bachrach et al., *Personality and Patterns of Facebook Usage*, PROC. 4TH ANN. ACM WEB SCI. CONF. 24, 24 (2012).

61. Sofia Ranchordás, *Nudging Citizens Through Technology in Smart Cities*, 34 INT'L REV. L. COMPUTERS & TECH. 254, 260 (2020) ("Nowadays, policymakers do not need to guess citizens' preferences. Instead, complex networks of millions of 'smart devices' (IoT) in constant communication, will give them access to solid information on what drives citizens with a greater degree of precision than any behavioral experiment would.").

62. ANDREW MCSTAY, EMOTIONAL AI: THE RISE OF EMPATHIC MEDIA 56 (2018) (describing a Facebook patent for using cell phone cameras to capture the user's "facial expressions, emotions and reactions"); FRANK PASQUALE, NEW LAWS OF ROBOTICS 130 (2020) ("AI can read sorrow, joy, disgust, and many other feelings from video of faces.").

am considering a payday loan and predicts, based on my search history, past check writing practices, and current bills, that I am borrowing the money to pay off my electric bill, it might disclose alternate payment plans the utility company offers. Alternatively, it might predict that I am going to ask my current partner to marry me based on my online research about engagement rings. It might then disclose the curious fact that "[m]en who spend between $2,000 and $4,000 on engagement rings are more likely to divorce than men who spend between $500 [and] $2,000."[63]

Once AI becomes able to predict our plans, preferences, and predicaments this well, along with many other aspects of the relevant decision environment, disclosure becomes an outdated paradigm. This is because an AI this knowledgeable and sophisticated has a superior option: Advice. The next Part offers a detailed defense of this claim.

II. ADVANTAGES OF THE ADVICE PARADIGM

Advice is quintessentially a recommendation about some course of action. For example, advice includes recommendations to do something, to refrain from doing something, or to use particular decision procedures.[64] Accordingly, we would all recognize the following as advice: "Read this article," "Don't send money to someone you just met online," and "Get a second opinion before agreeing to surgery." Disclosures offer facts that people must still expend time and energy processing to come to a decision. Advice cuts to the chase and offers a shortcut to the effortful processing that disclosures—even personalized ones—may still require.

The usefulness of advice is only partially about its content. The way that natural advice is packaged exerts a strong influence on whether it will ultimately affect behavior. Advice is often accompanied by words and actions that promote the advisee's self-efficacy, reinforce the advisee's self-control, and preemptively disarm knee-jerk psychological defensiveness. Advice can also be bundled with precommitment mechanisms, monitoring, and reminders. These are all part of the packaging that surrounds the recommendation that natural advice provides. This Article will refer to the dichotomy

63. Caitlin Davis Hornback, *Making Marriage: Some Key Ingredients for a More Successful Union*, STEPHEN F. AUSTIN STATE UNIV.: SCHOLARWORKS, https://scholarworks.sfasu.edu/cgi/viewcontent.cgi?article=1253&context=urc [https://perma.cc/44LS-9FLS] (last visited May 12, 2021).

64. Reeshad S. Dalal & Silvia Bonaccio, *What Types of Advice Do Decision-Makers Prefer?*, 112 ORG. BEH. & HUM. DECISION PROCESSES 11, 11-12 (2010).

between content and packaging as a useful shorthand, even though the distinction is fuzzy at times.[65]

Consider how a payday borrower might interact with natural advice. We might imagine the following conversation:

> *Friend: "You know, lots of people rollover those loans, and end up paying way more than they expected. I heard Alex got into trouble with them last year."*
>
> *Sam: "That won't happen to me! I always pay my bills on time."*
>
> *Friend: "Like last month's cable bill?"*
>
> *Sam: "That was different because..."*
>
> *Friend: "Seriously? Well, what about that gym membership that you got and then never visited? Look, even responsible people can get caught up in the moment or fall behind on bills. What if something happens this month and you can't pay the loan back—you know, like if your kid needs a new band uniform. You do a great job managing money, but this payday loan seems too dangerous. At least ask your sister for help first."*
>
> *Sam: "Yeah, I guess I should do that."*
>
> *Friend: "And promise me that if you get a payday loan, get only what you need. Don't borrow $500 when you could get away with only borrowing $200. Borrowing too much is just dumb."*

Notice how the conversation includes debiasing efforts, affirmations, informal precommitment devices, disclosures, and advice. It also backs up that advice with dynamic reason-giving.

The following subsections explore the advantages of advice in detail. They do so in the context of what I call *idealized* natural advice. That is, I assume that the advised course of action would be welfare-enhancing,[66] and that the packaging increases the likelihood that the advice will be heeded. This is not an empirical assumption about advice in the real world. Sometimes we all get and give bad advice. I use idealized natural advice to illustrate the potential of the advice paradigm. As I will argue later, many elements of idealized natural

65. For example, written advice will have to use certain words and certain font sizes, each of which might influence readers. All content must be packaged to some degree, and so they are not fully separable. For a similar point on the inevitability of subtle influence, see Christine Jolls et. al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471, 1534 (1998).

66. Compared, that is, to the course of action that the advisee would have taken absent the particular piece of advice.

advice can be incorporated into AI advice. Others can only be partially incorporated.

In a companion piece, I analyze natural advice without this idealization assumption.[67] That article brings together disparate lines of research to argue that natural advice will often be a useful resource for regulators to tap into.[68] In fact, there are benefits to a hybrid system that takes advantage of both AI advice and natural advice.[69] The burden of this Article is to show the immense potential of AI advice.

A. Cutting to the Chase: Sidestepping Literacy and Numeracy, Dealing with Decision Fatigue, and Helping Make Difficult Tradeoffs

Big data disclosures, as currently envisioned, are still just disclosures.[70] They present information to people as inputs to allow them to make an active (and now better informed) choice. But there are severe limits to this paradigm. Personalization can simplify disclosures, but only at the expense of communicating less information.[71] It can make some features of the disclosure more salient, and others less salient. This is certainly helpful in many circumstances,[72] but still requires the recipient to read the information and expend effort to process it accurately.

Literacy and numeracy each limit the effect of even the most personalized disclosures. Forty million Americans are functionally illiterate.[73] Only 20 percent of people can divide two numbers to figure out the cost per ounce of a jar of peanut butter.[74] About 40 percent of people cannot convert a proportion into a percentage.[75] In complex situations, mandatory disclosure cannot simplify the information enough to accommodate the limits of our literacy and numeracy. Complex products are, well, complex, and simplification has limits.

Some people have recommended financial education as a necessary corollary to mandated disclosure.[76] But it too has limits.[77] Financial

67. *See generally* Sean Hannon Williams, *Advice,* UTAH L. REV. (forthcoming) (manuscript on file with author).

68. *See id.* at Part II.

69. *See id.* at Part III.

70. *See* Porat & Strahilevitz, *supra* note 2, at 1445, 1474 (discussing examples of personalized disclosures such as disclosing, for different people, different possible medical side effects, different allergy information, and different terms within a proposed contract).

71. BEN-SHAHAR & SCHNEIDER, *supra* note 7, at 7.

72. Porat & Strahilevitz, *supra* note 2, at 1472 (arguing that personalized disclosures can mitigate information overload by making disclosures "shorter and more relevant").

73. Ben-Shahar & Schneider, *supra* note 7, at 711.

74. Lauren Willis, *Against Financial-Literacy Education*, 94 IOWA L. REV. 197, 222 (2008).

75. Ben-Shahar & Schneider, *supra* note 7, at 712.

76. Willis, *supra* note 74, at 199-200.

77. *Id.* at 201.

education cannot bridge the large gap between our baseline numeracy and the skills that would be required to understand many of today's financial products.[78] If consumers cannot figure out the price per ounce of peanut butter, they are unlikely to be able to do the calculations necessary to understand things like the time value of money. Ample evidence supports this. In one study, approximately 82 percent of people approaching retirement could not answer the following question correctly: "Let's say you have 200 dollars in a savings account. The account earns 10 percent interest per year. How much would you have in the account at the end of two years?"[79] Compare this to the level of skill necessary to make an informed choice about retirement savings, which requires knowledge of compound interest, mortality tables, the relative return on stocks versus bonds, the cumulative effect of a 1 percent fee versus a .5 percent fee, etc. In the face of these realities, the paradigm of disclosure is fundamentally limited, regardless of how personalized it is.

Even if people had the mathematical ability to make these and other computations for any single decision, the sheer number of decisions that they face makes careful consideration a Sisyphean task. Decision fatigue sets in and results in less-effortful processing. People look for short cuts. Toward the end of a long day, doctors prescribe more unnecessary antibiotics and wash their hands less often,[80] helpline nurses recommend seeing another health care professional rather than more definitively answering the patients' questions,[81] and car buyers are more likely to accept the default option that a salesperson provides.[82]

Additionally, personalized disclosures may not work as well for decisions that require difficult tradeoffs. Consider payday loans. They require a judgment about the likelihood of repaying, the harms of not repaying on time, the harms of not having the added cash now, etc. A personalized disclosure might rely on the fact that this particular person has a particular credit card. It might say: "This loan has an effective APR of 400 percent, compared to your Discover credit card which is currently charging 15.5 percent for a cash advance." This is good and useful information, but notice several limitations. It does not tell the person how to get the cash advance, how quickly they would get the money, or whether they could get the same amount of cash from

78. *Id.*; *see* BEN-SHAHAR & SCHNEIDER, *supra* note 7, at 46-47.

79. Annamaria Lusardi & Olivia S. Mitchelli, *Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education*, 42 BUS. ECON. 35, 37 (2007). The answer is $242.

80. J.L. Allan et al., *Clinical Decisions and Time Since Rest Break: An Analysis of Decision Fatigue in Nurses*, 38 HEALTH PSYCH. 318, 319 (2019).

81. *Id.* at 321.

82. *See* Jonathan Levav et al., *Order in Product Customization Decisions: Evidence from Field Experiments*, 118 J. POL. ECON. 274, 296 (2010).

both sources. The personalized disclosure could offer this extra information and could perhaps structure it to minimize its impact on the salience of the first and more important sentence. But the decision will only be easy when one option clearly dominates the other. When tradeoffs have to be made—perhaps it takes five days to get $500 from the credit card company, but only five minutes to get $1,000 from the payday lender—it is not clear that the disclosure will be helpful. Even if an attentive person could sift through the disclosures and conclude that one choice dominates the others, the process of doing so is itself taxing and stress-inducing.

Advice is a better paradigm. Advice provides an answer to the questions raised by difficult tradeoffs. To use the payday loan example above, the advice might be: "Don't get the payday loan. The credit card is a much better deal, and you can afford to wait a few days for the money." The goal of advice is not to present the information relevant to making an active choice. Rather, the goal is to predict what is best for the person and tell them what they should do. Advice circumvents literacy and numeracy. "Don't get the payday loan" is easy to understand and requires no numerical ability to comprehend. Advice also alleviates decision fatigue and helps people make difficult tradeoffs because it simply offers you the answer.[83]

Trends in the medical field provide additional reasons to favor advice over disclosure. Medical care today focuses on patient autonomy.[84] Doctors provide patients with detailed information, and patients decide for themselves. This works for low stakes decisions, but when the stakes are high, patients generally do not want autonomy.[85] They do not want the responsibility and potential regret of making these difficult decisions.[86] At the very least, they want more than information from their doctor. More specifically, they want *advice*.[87] They ask: "What would you do, Doc?" Studies of patient

83. Of course, people have to decide to follow the advice. But advice gives them someone else to blame if things go wrong, taking some of the pressure off of a choice to follow advice. Notice also that AI advice requires a great deal of knowledge to make difficult tradeoffs. In order for the advice to help you make the difficult tradeoff above, the AI must know why you need the money or at least how the delay in receiving it might affect you. Given our increasingly data-generating lives, the AI may well have this information.

84. Tracey King Schaller & Naresh K. Malhotra, *Affective and Cognitive Components of Attitudes in High-Stakes Decisions: An Application of the Theory of Planned Behavior to Hormone Replacement Therapy Use*, 32 PSYCH. & MARKETING 678, 690 (2015).

85. Mellina da Silva Terres et al., *Antecedents of the Client's Trust in Low- versus High-Consequence Decisions*, 29 J. SERVICES MARKETING 26, 27 (2015).

86. *See* Noel T. Brewer at al., *Anticipated Regret and Health Behavior: A Meta-Analysis*, 35 HEALTH PSYCH. 1264, 1265, 1269-70 (2016); Wendy Levinson et al., *Not All Patients Want to Participate in Decision Making: A National Study of Public Preferences*, 20 J. GEN. INTERNAL MED. 531, 532 (2005) ("Fifty-two percent preferred to leave final decisions to their physicians.").

87. William M. Strull et al., *Do Patients Want to Participate in Medical Decision Making?*, 252 JAMA 2990, 2990-94 (1984); Karen A. Scherr et al., *Physician*

preferences reveal that between 50 percent and 80 percent of patients actually want their doctor to decide for them.[88] In a world where we do not allow people to decide for one another, asking the doctor what they would do is likely to be a close second to just having them decide for you.

B. Reason-Giving: Static and Dynamic

One common addendum to a piece of advice is the reasons that support it. Recall the illustration of Sam's payday loan. His friend's advice might include anecdotes about others ("Alex got an auto title loan, and they took his car!") or information about Sam himself ("Remember last time when you borrowed money from Brenda, and you had a hard time paying it back?"). It might also include explanations of the underlying logic ("At the end of the month, you'll owe less money if you borrow using the credit card; let me explain . . ."). And of course, the friend would be attempting to personalize those explanations to account for Sam's literacy, numeracy, and attention span.

Reason-giving makes advice more potent. It increases the advisee's trust in the advisor and the perceived quality of the advice.[89] It also allows the advisee to verify that the advice is relevant and to ensure that the advisor is not misunderstanding the situation.

Natural advice is also dynamic rather than static. The advisee might push back or ask questions, and because this is a natural conversation between two people, the advisor can answer. For example, when the advisee says, "This time will be different!" the advisor can say, "Yeah, that's what you said last time." Advisors also respond to changes in the advisee. If someone starts getting angry at the advice, the advisor might seek to calm them down before continuing the conversation. If someone starts to resist the advice, the advisor might escalate the severity of the warnings, change the subject

---

Recommendations Trump Patient Preferences in Prostate Cancer Treatment Decisions, 37 MED. DECISION MAKING 56, 65-66 (2016) (finding that patients' pre-visit preferences, values, and attitudes were not predictive of their actual treatments, but that their doctors' recommendations were).

88. Levinson, *supra* note 86, at 532 (52%); Neeraj K. Arora & Colleen A. McHorney, *Patient Preferences for Medical Decision Making: Who Really Wants to Participate?* 38 MED. CARE 335, 336 (2000) (reporting that 69% of patients wanted to leave the final decision to their doctor); John D. Lantos, *Do Patients Want to Participate in Decisions About Their Own Medical Care?*, 15 AM. J. BIOETHICS 1, 1 (2015) (80%); Terres et al., *supra* note 85, at 27 (finding that, as the decision becomes more high consequence, people's desire for autonomy and responsibility is reduced and they start experiencing the negative emotions stemming from having to choose).

89. Y. Yi Mun et al., *Untangling the Antecedents of Initial Trust in Web-Based Health Information: The Roles of Argument Quality, Source Expertise, and User Perceptions of Information Quality and Risk*, 55 DECISION SUPPORT SYS. 284, 291 (2013).

for a time, or switch to some other line of argument. These are not techniques that only the savviest advisors use; they come naturally to many people.

C. *Overcoming Motivated Reasoning: A Spoonful of Sugar Helps the Medicine Go Down*

Motivated reasoning significantly impairs the effectiveness of warnings and other disclosures.[90] People resist bad news. When people are given the opportunity to read about risk factors for a disease, they will tend to read more about those factors that offer them good news. For example, smokers who have low cholesterol would disproportionately read about the links between cholesterol and heart disease and avoid reading about the links between smoking and heart disease.[91] Even if they read about the factors that offer them bad news, they would not retain as much of this negative information and would be more likely to challenge it.[92] This resistance to bad news has been consistently found in a host of contexts,[93] including financial planning,[94] education,[95] politics,[96] and health.[97]

Although people resist bad news, friends and family can deliver the necessary information in ways that make people far more likely to listen to it. An advisor who knows that she is recommending a hard or awkward road may say something encouraging. In advising Sam about

---

90. Sean Hannon Williams, *Sticky Expectations: Responses to Persistent Over-Optimism in Marriage, Employment Contracts, and Credit Card Use*, 84 NOTRE DAME L. REV. 733, 748 (2009); Nicholas Epley & Thomas Gilovich, *The Mechanics of Motivated Reasoning*, 30 J. ECON. PERSP. 133, 136 (2016) ("Most people do not reason like impartial judges, but instead recruit evidence like attorneys, looking for evidence that supports a desired belief while trying to steer clear of evidence that refutes it.").

91. *See* Nathan Radcliffe & William M.P. Klein, *Dispositional, Unrealistic, and Comparative Optimism: Differential Relations with the Knowledge and Processing of Risk Information and Beliefs About Personal Risk*, 28 PERS. & SOC. PSYCHOL. BULL. 836, 843 (2002) (evaluating risk factors for one's own health).

92. *See id.*; Peter H. Ditto & David F. Lopez, *Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions*, 63 J. PERS. & SOC. PSYCHOL. 568, 568 (1992).

93. Constantine Sedikides & Jeffrey D. Green, *What I Don't Recall Can't Hurt Me: Information Negativity Versus Information Inconsistency as Determinants of Memorial Self-Defense*, 22 SOC. COGNITION 4, 22 (2004).

94. Lauren Willis, *Against Financial-Literacy Education*, 94 IOWA L. REV. 197, 235-36 (2008); JAMES M. LACKO & JANIS K. PAPPALARDO, BUREAU OF ECONS. FTC, STAFF REPORT, THE EFFECT OF MORTGAGE BROKER COMPENSATION DISCLOSURES ON CONSUMERS AND COMPETITION: A CONTROLLED EXPERIMENT, 28-29 (2004) (noting that some borrowers refused to believe that their mortgage had a prepayment penalty even after it was pointed out to them).

95. Erik Helzer & David Dunning, *Why and When Peer Prediction is Superior to Self-Prediction: The Weight Given to Future Aspiration versus Past Achievement*, 103 J. PERS. SOC. PSYCHOL. 38, 49 (2012).

96. Martin Baekgaard et al., *The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians*, 49 BRITISH J. POL. SCI. 1117, 1117 (2017).

97. Radcliffe & Klein, *supra* note 91, at 843.

whether to get a payday loan, the friend said: "You are great with money." A large body of research shows that affirming someone's self-worth reduces defensiveness and helps people read and digest threatening information.[98] As mentioned above, patients routinely resist negative information about their health and actively avoid disclosures that would help them learn more about the risk factors they possess.[99] But this trend can be reversed by affirming the patient's self-worth and self-efficacy in certain ways.[100] After this seemingly mild intervention, patients actively seek out more information about their potential health problems.[101] They do so because they now have more emotional resources to interact constructively with bad news.[102] These effects are not limited to the health context. In a sadly familiar story, political partisans routinely reject information or arguments that challenge their preferred policy outcome.[103] But these partisan effects disappear when people's self-worth is bolstered.[104]

Relatedly, affirmations help people manage stress more effectively, as indicated by cortisol and epinephrine levels—both stress hormones.[105] This in turn reduces distortions created by stress. In one terrifying study, people had their head immobilized and then experimenters placed a large tarantula near their face.[106] The resulting stress distorted their estimate of how far away the large

98. Geoffrey L. Cohen & David K. Sherman, *The Psychology of Change: Self-Affirmation and Social Psychological Intervention*, 65 ANN. REV. PSYCHOL. 333, 339, 347-48 (collecting studies) [hereinafter Cohen & Sherman, *The Psychology of Change*]. *See generally* William M.P. Klein & Peter R. Harris, *Self-Affirmation Enhances Attentional Bias Toward Threatening Components of a Persuasive Message*, 20 PSYCHOL. SCI. 1463 (2009); David K. Sherman & Geoffrey L. Cohen, *Accepting Threatening Information: Self-Affirmation and the Reduction of Defensive Biases*, 11 PSYCHOL. SCI. 119 (2002).

99. *See* Cohen & Sherman, *The Psychology of Change*, *supra* note 98, at 339.

100. *Id.*; *see also* Guido Van Koningsbruggen, Enny Das & David R. Roskos-Ewoldsen, *How Self-Affirmation Reduces Defensive Processing of Threatening Health Information: Evidence at the Implicit Level*, 28 HEALTH PSYCHOL. 563, 563 (2009).

101. *See* Van Koningsbruggen & Roskos-Ewoldsen, *supra* note 100, at 563.

102. *See id.*

103. Peter H. Ditto et al., *At Least Bias is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives*, 14 PERS. PSYCHOL. SCI. 273, 273 (2019) (finding these effects for both liberals and conservatives).

104. Kevin R. Binning et al., *Seeing the Other Side: Reducing Political Partisanship via Self-Affirmation in the 2008 Presidential Election*, 10 ANALYSES SOC. ISSUES & PUB. POL'Y 276, 285-86 (2010) (partisanship and Obama's election). *See generally* Geoffrey L. Cohen et al., *Bridging the Partisan Divide: Self-Affirmation Reduces Ideological Closed-Mindedness and Inflexibility in Negotiation*, 93 J. PERS. & SOC. PSYCHOL. 415 (2007) (partisanship and abortion).

105. Cohen & Sherman, *The Psychology of Change*, *supra* note 98, at 339-40, 350.

106. *Id.* at 339.

spider was.[107] But affirmations entirely eliminated this visceral distortion.[108]

The basic explanation for these surprisingly robust effects is that affirming one's core values helps people put various threats into context.[109] Information that problematizes one's political commitments is threatening, but it is not anyone's entire life. When, for example, people first reflect on friends and family, they see many threats as smaller and less significant.[110] Those threats are isolated events within a larger and more secure self-concept.[111] Accordingly, they respond to threats in a calmer and more productive way.[112] Friends and family are likely to understand when advice is threatening and are likely to be able to mitigate that threat through affirmations and other emotional support.

Lauren Willis identified various strong emotions like dread and anxiety as barriers to translating financial education into better financial choices.[113] She mocked faith in financial education by arguing that most people would need substantial psychotherapy to overcome these emotional issues:

> Some biases that impede good financial decision-making . . . may require individualized psychotherapy. For an individual whose self-worth is tied to materially keeping up with his or her social group, but whose spending to reach this goal outstrips financial resources, a lack of financial information, skills, confidence, or motivation is not the problem. But psychotherapy is not what policymakers mean when they talk about financial education.[114]

Advice might be a shortcut to the lengthy therapy that Willis jokingly recommends, in part, because natural advice can offer the emotional support necessary to reinforce one's self worth and otherwise reduce the effects of heightened emotions.

---

107. *Id.*

108. *Id.*; *see also* Loreal Shea & E.J. Masicampo, *Self-Affirmation Counteracts the Effects of Burdens on Judgments of Distance*, 50 J. EXPERIMENTAL SOC. PSYCHOL. 105, 106-107 (2014) (finding that people perceived a hill as steeper than it actually was, unless they were affirmed first); J. David Creswell et al., *Affirmation of Personal Values Buffers Neuroendocrine and Psychological Stress Responses*, 16 PSYCHOL. SCI. 846, 847-49 (2005) (finding that people did not experience heightened stress from public speaking, as long as they were affirmed first).

109. Cohen & Sherman, *The Psychology of Change*, *supra* note 98, at 339.

110. *Id.*

111. *Id.* at 339, 342.

112. *Id.* at 339.

113. Willis, *supra* note 74, at 252-53.

114. *Id.*

D. Enhancing Self-Control and Self-Efficacy

Self-control problems are likely involved in many decisions, yet disclosures do nothing to address them.[115] Credit card borrowing is the quintessential example. Credit card companies have arguably designed their product to distance the pain of paying from the joy of getting, thereby exacerbating self-control problems.[116] Disclosures, regardless of how personalized they are, do not address these self-control problems. Someone with a self-control problem is likely to make a short-sighted choice even if a perfectly designed disclosure has revealed all of the relevant information.[117]

Natural advice, in contrast, has the potential to reduce self-control problems. The affirmations that can accompany natural advice not only reduce motivated reasoning, but they also increase self-control.[118] For example, not only do affirmed people seek out more information about threatening health conditions, but they are also more likely to alter their behavior.[119] Affirmations help people lose weight,[120] quit smoking,[121] and to continue to take their medication.[122] The effects of affirmations are large and enduring. For example, one year after being affirmed, patients were still significantly more likely to be taking their medication.[123] A weight loss study found that, more than two and a half months after being affirmed, subjects were more than six pounds lighter than the subjects who were not affirmed.[124]

Encouragement, like affirmations, can also help people take the harder but more rewarding road. Self-talk—the name given to saying to yourself, "I can do this" during challenges—has been extensively

115. Bubb & Pildes, *supra* note 14, at 1642; John Gathergood, *Self-Control, Financial Literacy and Consumer Over-Indebtedness*, 33 J. ECON. PSYCHOL. 590, 591 (2012) (finding that "over-indebtedness, measured both as delinquency on repayments and self-reported financial distress, occurs disproportionately among individuals who report self-control problems").

116. Arguably, mandated disclosure and financial education exacerbate the problem. By highlighting how important the underlying decisions are, they might induce more stress, which tends to reduce willpower. Similarly, taxing cognitive resources—by engaging seriously with the underlying financial math—tends to leave fewer resources for self-control. Lauren E. Willis, *The Financial Education Fallacy*, 101 AM. ECON. REV. 429, 430 (2011); Willis, *supra* note 74, at 227.

117. *See* Bar-Gill, *supra* note 59, at 1396 n.112.

118. Cohen & Sherman, *The Psychology of Change*, *supra* note 98, at 339-40.

119. *Id.* at 339.

120. *Id.* at 351.

121. *Id.* at 347-49 (collecting studies).

122. *Id.* at 351.

123. Gbenga O. Ogedegbe et al., *A Randomized Controlled Trial of Positive-Affect Intervention and Medication Adherence in Hypertensive African Americans*, 172 ARCHIVES INTERNAL MED. 322, 324 (2012).

124. Christine Logel & Geoffrey L. Cohen, *The Role of the Self in Physical Health: Testing the Effect of a Values-Affirmation Intervention on Weight Loss*, 23 PSYCHOL. SCI. 53, 54 (2012) (finding that, on average, affirmed subjects lost 3.41 pounds and non-affirmed subjects gained 2.76 pounds).

studied by sports psychologists. The consistent finding is that self-talk boosts athletic performance and helps people persevere in the face of challenges.[125] It helps people swim and run faster during real competitions and has been shown to improve a host of physical performance measurements in more controlled environments.[126] These encouragements are likely to be more effective when they come from others. In a recent study, cyclists were asked to use self-talk during a time trial on a stationary bike.[127] The researchers found that saying things like "*You* can do this" improved performance significantly more than saying "*I* can do this."[128] Just as seeing a picture of a loved one can help you endure more pain,[129] even a subtle cue that others believe in you can help you focus on your goals despite the physical pain.

Bolstering self-efficacy (the feeling that you can accomplish your goals) is often central to motivating action. In the context of health, much has been written about how to effectively motivate healthy behaviors. Using various "persuasive technologies," researchers have tried to make people eat better, exercise more, etc.[130] Importantly, these efforts are not just focused on information, but rather also on providing people with a sense of self-efficacy.[131] Several studies have found that, at least for western cultures, *self-efficacy is the most*

125. Antonis Hatzigeorgiadis et al., *Self-Talk and Sports Performance: A Meta-Analysis*, 6 PERS. PSYCHOL. SCI. 348, 349 (2011); Antonis Hatzigeorgiadis et al., *Self-Talk and Competitive Sport Performance*, 26 J. APPLIED SPORT PSYCHOL. 82, 83 (2014).

126. Hatzigeorgiadis et al., *supra* note 125, at 83, 89 (swimming and running); Adrian Hase et al., *The Influence of Self-Talk on Challenge and Threat States and Performance*, 45 PSYCHOL. SPORT & EXERCISE 101550, Section 3.2.H2. (2019) (darts); Yannis Theodorakis et al., *Cognitions: Self-Talk and Performance*, *in* THE OXFORD HANDBOOK OF SPORT AND PERFORMANCE PSYCHOLOGY 191, 197 (Shane M. Murphey ed., (2012) (tennis, basketball).

127. James Hardy et al., *To Me, to You: How You Say Things Matters for Endurance Performance*, 37 J. SPORTS SCI. 2122, 2123-24 (2019).

128. *Id.* at 2123, 2127 (emphasis added).

129. *See, e.g.*, Sarah L. Master et al., *A Picture's Worth: Partner Photographs Reduce Experimentally Induced Pain*, 20 PSYCHOL. SCI. 1316, 1317 (2009) (finding that "simply viewing a loved one's picture can have pain-attenuating effects"); Naomi I. Eisenberger et al., *Attachment Figures Activate a Safety Signal-Related Neural Region and Reduce Pain Experience*, 108 PNAS 11721, 11723 (2011) (same); Jarred Younger et al., *Viewing Pictures of a Romantic Partner Reduces Experimental Pain: Involvement of Neural Reward Systems*, 5 PLOS ONE e13309, 1 (2010) (finding that viewing photos of significant others triggers reward centers of brain, which has an effect of reducing pain).

130. Rita Orji & Regan L. Mandryk, *Developing Culturally Relevant Design Guidelines for Encouraging Healthy Eating Behavior*, 72 INT. J. HUMAN-COMPUTER STUD. 207, 210 (2014).

131. *Id.* at 211-12, 218. Some researchers have done so through video games that help dynamically illustrate ways to overcome various barriers and provide direct practice and feedback to support self-efficacy. *Id.* at 212.

important determinant of behavioral change.[132] Standard information disclosures do not help self-efficacy,[133] natural advice can.

Finally, advice is often bundled with reminders to keep the long-term goal more salient. Someone might say: "I know that look in your eye. Don't do it!" Reminders of one's long-term goals can help people overcome moments of weakness, or just help forgetful people remember, for example, to take their medication. [134]

E. Providing Opportunities for Precommitment

Many advice-seeking conversations will trigger informal, but important, precommitments.[135] One could easily imagine a friend advising you to pay off your credit card balance if you have enough money in your checking account to do so.[136] In the course of this conversation, you may well respond: "Yeah, I guess I should do that." This is a very mild precommitment, but it is not without weight. When your friend checks back in the next month, you would likely feel embarrassed if you ignored their advice. Just casually indicating that you intend to sign up for an employee retirement program, or that you will get a second opinion about a medical procedure, makes it more likely that you will do so.[137]

F. Maintaining Transparency: Comparing Advice to Nudges

Advice is transparent rather than manipulative. Advice openly and obviously attempts to influence your choice, and it does so through rational persuasion.[138] These features of the advice paradigm set it apart from some classic nudges.

A nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or

---

132. *See id.* at 218.

133. Of course, big data disclosures could also promote self-efficacy by being more explicit about the concrete steps one can take to achieve a certain result. Instead of saying, "There may be alternative sources for you to borrow from, including credit cards," an AI may be able to say, "If you call Discover Card, they can provide you with a cash advance in about 10 minutes, click here to call them now." But when the best course of action is more difficult than merely clicking on a link, messages should move beyond factual disclosure to more explicitly target self-efficacy.

134. Jonathan E. Mayer & Paul Fontelo, *Meta-Analysis on the Effect of Text Message Reminders for HIV-Related Compliance*, 29 AIDS CARE 409, 411 (2017).

135. Of course, these informal self-control mechanisms could be achieved in the absence of advice, but advice usefully bundles them with other benefits.

136. David B. Gross & Nicholas S. Souleles, *Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data*, 117 Q. J. ECON. 149, 178 (2002) (finding that many consumers carry both credit card debt and liquid assets that would allow them to pay down the debt or borrow at lower interest rates).

137. *See infra* Section III.B.3.

138. For a potential caveat, *see infra* note 231 and accompanying text.

significantly changing their economic incentives."[139] Under this capacious definition, disclosures are nudges, reminders are nudges, and advice is a nudge.[140] But there exists a classic set of nudges that are a subset of this expansive definition. Those seek to influence behavior by taking advantage of biases.[141] Classic examples include automatically opting people into a retirement savings plan, rearranging the order of food in a cafeteria so that people fill their tray with fruits and vegetables before getting cake, or informing people that their peers do something in order to capitalize on social pressures to conform.[142]

There are some reasons to think that natural advice will compete favorably with many nudges. A large body of evidence shows that people follow good advice and discount bad advice.[143] This literature generally illustrates the great potential of natural advice to improve welfare. However, there is a dearth of evidence directly comparing the efficacy of advice to nudges. This is not a surprise given that this is the first Article to take advice seriously as a regulatory tool.

Given this gap in the current data, this subsection sets welfare aside and focuses on the advantages of advice from the perspective of autonomy. Here, advice is importantly different from classic examples of nudging.

Critics of nudging often equate it with manipulation.[144] This objection is rooted in the claim that nudges are covert rather than transparent.[145] Covert attempts to alter choice reduce autonomy.[146]

---

139. THALER & SUNSTEIN, *supra* note 11, at 6.

140. Cass R. Sunstein, *People Prefer System 2 Nudges (Kind of)*, 66 DUKE L.J. 121, 124 (2016) ("A reminder is a nudge; so is a warning. A GPS device nudges; a default rule nudges. Disclosure of relevant information . . . counts as a nudge. A recommendation is a nudge."); *but see* Daniel M. Hausman & Brynn Welch, *Debate: To Nudge or Not to Nudge*, 18 J. POL. PHIL. 123, 127 (2010) (arguing that advice should not be considered a nudge).

141. Sunstein, *supra* note 140, at 125, 128 ("Noneducative nudges [or system 1 nudges] include default rules and strategic decisions about how items are ordered (as on a menu or at a cafeteria); these are designed to preserve freedom of choice without necessarily increasing individual agency.").

142. *See* CASS R. SUNSTEIN, THE ETHICS OF INFLUENCE: GOVERNMENT IN THE AGE OF BEHAVIORAL SCIENCE 26-27 (2016).

143. *See generally* Williams, *supra* note 67.

144. Pelle Guldborg Hansen & Andreas Maaløe Jespersen, *Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy*, 4 EUR J. RISK REG. 3, 5 (2013) ("A persistent and central tenet in the political and normative criticism has been the claim that nudging works by 'manipulating people's choices.'").

145. *Id.* at 5, 17; SUNSTEIN, *supra* note 142, at 105 (discussing whether failing to disclose the power of a default rule make its use manipulative).

146. Hendrik Bruns et al., *Can Nudges Be Transparent and Yet Effective?*, 65 J. ECON. PSYCHOL. 41, 42 (2018); T. J. Kasperbauer, *The Permissibility of Nudging for Sustainable Energy Consumption*, 111 ENERGY POL'Y. 52, 52 (2017) ("[One] objection is that nudges reduce human autonomy and agency. In short, these objections claim that nudges inappropriately interfere with people's decisions, in a way that reduces people's ability to live their lives the way they want to live them.").

Subliminal messages are the quintessential example.[147] This level of manipulation is objectionable regardless of whether the messenger is a government or a for-profit company.[148] Other manipulations are judged differently depending on the messenger; governments have heightened duties to be transparent so that citizens can monitor and control government action.[149] Is it manipulative for a government-run cafeteria to shrink the size of lunch plates to reduce calorie intake? This might be a close call. Some scholars classify it as manipulation, while others think it is sufficiently transparent because a "watchful person" would notice the plate size and could surmise, or ask sufficient questions to understand, the nudge and its purpose.[150] Big data nudges are likely even more objectionable on transparency grounds because most people cannot understand the complex algorithm that led to the nudge.[151] This thwarts attempts by citizens to understand the ways that they are being influenced, to understand the justifications for that influence, and to monitor substantive nudging decisions that governments make to influence them.[152] Many people agree that some

147. SUNSTEIN, *supra* note 142, at 82; THALER & SUNSTEIN, *supra* note 11, at 245-46 (arguing that subliminal advertising is objectionable even if disclosed beforehand because its influence would be "invisible and thus impossible to monitor").

148. ZACK LYNCH & BYRON LAURSEN, THE NEURO REVOLUTION: HOW BRAIN SCIENCE IS CHANGING OUR WORLD 53 (2009) (quoting the FCC as concluding that "[s]ubliminal advertising is intended to be deceptive, and is contrary to the public interest").

149. SUNSTEIN, *supra* note 142, at 96-97 (arguing that the wrongfulness of manipulation varies depending on whether the context is advertising or direct communications from the government to citizens).

150. Hansen & Jespersen, *supra* note 144, at 16-17 (manipulation); Andreas T. Schmidt, *The Power to Nudge*, 111 AM. POL. SCI. REV. 404, 411 (2017) (sufficiently transparent).

151. Karen Yeung, *'Hypernudge': Big Data as a Mode of Regulation by Design*, 20 INFO. COMM. & SOC. 118, 124 (2017) (noting that "algorithmic accountability" is difficult); Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 43 (2013) ("[W]hen big data analytics are increasingly being used to make decisions about individual people, those people have a right to know on what basis those decisions are made."). A related concern is that nudges, and especially big data nudges, might be too powerful. The more powerful a nudge is, the less effective transparency would be at allowing people a meaningful opportunity to resist the nudge. Henrik Skaug Sætra, *When Nudge Comes to Shove: Liberty and Nudging in the Era of Big Data*, TECH. SOC. 3, 6 (April 2019) (arguing that a nudge "becomes more problematic the more effective it is"); Hausman & Welch, *supra* note 140, at 132 ("It seems to us that what bothers Thaler and Sunstein about the hypothetical case of efficacious subliminal advertising is the efficacy itself."); Yashar Saghai, *Salvaging the Concept of Nudge*, 39 J. MED. ETHICS 487, 492 (2013) (redefining nudges to exclude influences that are too powerful).

152. Hansen & Jespersen, *supra* note 144, at 17 ("[A] *transparent nudge* is defined as a nudge provided in such a way that the intention behind it, as well as the means by which behavioural change is pursued, could reasonably be expected to be transparent to the agent being nudged as a result of the intervention."); Hausman & Welch, *supra* note 140, at 135 ("One important way to protect against abuse and to respect autonomy is to make sure that the government actually inform people of efforts to shape their choices, not merely that it be able and willing to do so."); Schmidt, *supra* note 150, at 410 (discussing importance of democratic accountability); Hendrik Bruns et al., *supra* note 146, at 49 (discussing democratic legitimacy); *see* Porat & Strahilevitz, *supra* note 2, at 1459 (noting that personalized laws should use only broad categories like sex and age to ensure that courts can understand the underlying categories and assess them).

degree of transparency is required to make nudges ethically acceptable, but some disagree about the nature and degree of that transparency.[153]

These debates play out differently in the context of advice. Critics of nudging often create an ethical spectrum, where rational persuasion is the most autonomy-respecting and manipulation is far to the other side.[154] These critics often point to disclosures and advice as examples of unobjectionable rational persuasion.[155] Advice is unobjectionable for two reasons. First, advice is educative and seeks to engage the decider's rational system 2 processes.[156] Classic nudges are non-educative system 1 nudges[157] that some claim do "not preserve freedom of choice, because choice behavior is predominately influenced without the awareness of the individual."[158] Second, advice is (mostly) transparent.[159] Classic nudges might begin with a secret agenda and invisible influence, and it is only the watchful person who will discover them.[160] Advisors, by definition, identify themselves as others who are trying to influence your choice.[161] This is true regardless of whether

153. *See, e.g.*, Hausman & Welch, *supra* note 140, at 135 (discussing degrees of transparency); Sunstein, *supra* note 142, at 105 (discussing transparency about the psychological mechanisms that makes nudges effective); Sætra, *supra* note 151, at 7-8 (discussing transparency about the government's motive).

154. *See, e.g.*, Hausman & Welch, *supra* note 140, at 128; Saghai, *supra* note 151, at 492.

155. Hausman & Welch, *supra* note 140, at 128.

156. *See* Sunstein, *supra* note 140, at 124, 129 ("Educative nudges [or system 2 nudges] include disclosure requirements, reminders, and warnings, which are specifically designed to increase people's own powers of agency."); SUNSTEIN, *supra* note 142, at 34 (noting the system 2 nudges are less ethically controversial than system 1 nudges).

157. Sunstein, *supra* note 140, at 125, 128.

158. Yiling Lin et al., *Nudge: Concept, Effectiveness, and Ethics*, 39 BASIC & APPLIED SOC. PSYCHOL. 293, 301 (2017); *but see* Sunstein, *supra* note 140, at 134 (arguing that system 1 nudges may promote autonomy on net by freeing up cognitive space for other more important tasks).

159. To preserve these benefits, AI advice might be constrained to offer only truthful advice. Suppose the AI predicts that you would be better off if you did not go to the casino today. The AI also predicts that the best way to achieve this outcome is to provide you with an inaccurate disclosure by claiming that the casino is closed. It might alternatively predict that offering some distracting advice ("You should shop today because of the good sales going on now") would keep you out of the casino. Should the AI be allowed to lie or to mislead in these ways? This would potentially undermine the autonomy-advantages of advice. Accordingly, some form of no-lying constraint might be appropriate, at least absent an affirmative opt-in.

160. This secrecy is by no means a necessary feature of nudges. A company that alters a default rule for retirement savings could tell people that they are doing so and give honest reasons for their actions. Contrary to some early concerns, this openness does not always defeat the effectiveness of nudges. Luc Bovens, *The Ethics of Nudge*, *in* PREFERENCE CHANGE: APPROACHES FROM PHILOSOPHY, ECONOMICS AND PSYCHOLOGY 209 (Till Grüne-Yanoff & Sven O. Hansson eds., 2009) (arguing that the "psychological mechanisms that are exploited . . . work better in the dark"); Hendrik Bruns et al., *supra* note 146, at 48-49 (finding that nudges still work even when you explain the power of default rules and the purpose behind the choice of the default).

161. Hausman & Welch, *supra* note 140, at 135 (objecting to nudges because they do not inform people of the efforts to influence their choice).

the advice is solicited or not. Of course, some elements of natural advice are covert.[162] An advisor may use a soft smile to make you more susceptible to the advice or may subtly appeal to various strong emotions.[163] No communication operates solely through rational persuasion.[164] But advice flips the core and peripheral features of classic nudges. The core and obvious nature of advice is that someone else is trying to convince you that it is in your best interests to act in a certain way. The subtle features play out within this context, a context that would lead even a relatively oblivious person—not just a watchful one—to be on the lookout for potential manipulation.[165]

G. Summary

An analysis of natural advice reveals the powerful advantages of the advice paradigm over either disclosure or nudges. The content of advice sidesteps our limited numeracy and literacy, mitigates decision fatigue, and helps us deal with difficult tradeoffs. It also transparently relies on rational persuasion rather than invisible influence. The packaging of advice is also important. Reason-giving, reminders, and precommitment devices can all significantly enhance the effectiveness of advice. That packaging can also promote self-worth to overcome motivated reasoning and bolster self-efficacy to make behavioral

---

162. Saghai, *supra* note 151, at 492 n.ix ("[N]ot all forms of speech-based influences are reducible to rational persuasion (eg, deception, flattery, etc.).").

163. Advice is not incompatible with manipulation, and advisors who wish to incorporate direct pressure or nudges could do so. Someone whose friend says, "you'd be an idiot if you did X" would probably be disinclined to do it. If X turns out to be a bad decision, they will feel like an idiot. "I told you so" is often painful to hear, even if there is no threat that the friendship will end or deteriorate. Similarly, advice could incorporate social norm nudges, which rely on social comparisons to nudge choices in a particular direction. *See* Kasperbauer, *supra* note 146, at 53 (collecting studies on social norm nudges). For example, a friend might say, "even Jim figured out that the payday loan was a bad deal, and we all know you're smarter than Jim, right?" To the extent that these influences do not significantly affect the overall rational cost-benefit analysis of the decision, they would constitute classic nudges. SUNSTEIN, *supra* note 142, at 64 (noting that even strongly worded advice would not reduce autonomy because it would not impose substantial costs on deciding otherwise). Like other nudges, these features of advice might operate as debiasing tools or as more straightforward attempts to influence choice. Ryan Bubb, *TMI? Why the Optimal Architecture of Disclosure Remains TBD*, 113 MICH. L. REV. 1021, 1028-35 (2015).

164. Hausman & Welch, *supra* note 140, at 130 ("Our efforts to persuade may succeed because of the softness of our smile or our aura of authority rather than the soundness of our argument . . . .").

165. Rebecca Tushnet, *More Than a Feeling: Emotion and the First Amendment*, 127 HARV. L. REV. 2392, 2431 (2014) ("Compared to the effects of such nudges on autonomy, at least explicitly emotional appeals allow citizens to bring their 'persuasion knowledge' to bear and fight off the appeals.") Similar insights have led people to worry about online ads that look so much like non-advertising content that readers cannot distinguish them. *See, e.g.*, Irina Dykhne, *Persuasive or Deceptive? Native Advertising in Political Campaigns*, 91 S. CAL. L. REV. 339, 340 (2018).

change more likely. Perhaps this is why people crave advice rather than disclosures or nudges.[166]

III. LESSONS FOR AI ADVICE

This Part begins by highlighting one aspect of natural advice that AI advice may have the most difficulty mimicking, namely, the emotional connectedness between advisor and advisee. The remainder of the Part strikes a significantly more optimistic tone.

A. Can We Be Friends with an AI?

Several of the advantages of natural advice are likely to be more potent when the advisor is a friend or family member. For example, precommitments to friends are likely more effective than precommitments to strangers. Perhaps surprisingly, AI advice can, at least partially, mimic the relevant emotional connectedness.

AI is increasingly able to read and respond to emotions in appropriate ways.[167] Virtual assistants have already been developed that lightly mock people, or nag, insult, or praise them, or not-so-subtly compare them to others.[168] AI chatbots on Facebook rile us up,[169] and AI chatbots in call-centers calm us down.[170] Exercise apps praise and cajole.[171] People are more likely to form emotional attachments to

166. *See supra* notes 85-88 and accompanying text (suggesting a strong preference for advice over disclosure); Sunstein, *supra* note 140, at 156 (finding that "majorities will often prefer System 2 nudges [like factual disclosures] even when they believe that System 1 nudges are more effective").

167. MCSTAY, *supra* note 62, at 3, 7-11. Is a fist pump a sign of happiness or aggression? This is a very complicated question that people answer effortlessly based on context. But it is hard to train AI to do what we do so easily. *See* Steve Inskeep & Brenda Salinas, *Wimbledon, Steeped In Tradition, Embraces Artificial Intelligence*, NPR MORNING EDITION (July 9, 2019, 5:05 AM), https://www.npr.org/2019/07/09/739784008/wimbledon-steeped-in-tradition-embraces-artificial-intelligence [https://perma.cc/PZS2-A33U] (discussing the difficulty of training AI to understand gestures like fist pumps).

168. Deanna Pogorelc, *Awful But Funny: Talking Weight Loss App Dishes Out Insults to Help You Lose Weight*, MEDCITY NEWS (Feb 3, 2014, 7:00 AM), https://medcitynews.com/2014/02/dose-sarcasm-lose-weight-else/ [https://perma.cc/6CE8-9226]; *The Best Running Apps to Take on Your Workout*, RUNNER'S WORLD (May 19, 2019), https://www.runnersworld.com/gear/a20865699/best-running-apps/ [https://perma.cc/7FPP-FUS4] (discussing a running app that compares your exercise habits against nearby users to motivate you, and another that offers praise).

169. John Markoff, *Automated Pro-Trump Bots Overwhelmed Pro-Clinton Messages, Researchers Say*, N.Y. TIMES (Nov. 17, 2016), https://www.nytimes.com/2016/11/18/technology/automated-pro-trump-bots-overwhelmed-pro-clinton-messages-researchers-say.html [https://perma.cc/9UKZ-JCXT].

170. Friedman, *supra* note 29 (building into a customer service AI chatbot a "negative sentiment detector" in order to identify angry customers and respond with sympathetic statements).

171. *See supra* note 168; *see also* PASQUALE, *supra* note 62, at 6 ("[I]f thousands of trainers equipped themselves with Google Glass and recorded all their encounters, perhaps some divine database of grimaces and rolled eyes, injuries, and triumphs could dictate the optimal response to a miserable gym goer.").

AI if it looks and sounds like a person or animal.[172] Accordingly, virtual assistants are getting computer generated faces to go along with their now-disembodied voices.[173]

Future AIs might do better. Many movies attest to the possibility of falling in love or becoming friends with an AI,[174] and they may be on to something.[175] With the future power of AI and big data, we could customize various strategies for creating emotional connectedness. Your social media interactions may reveal your love language or just how you tend to communicate with good friends. The AI could then match these communication styles. It would flatter some and mock others. This will help the AI trigger at least some of the emotional responses that will make us more likely to hear and heed advice.

B. Layered Messages

As AI develops, it is possible that personalized advice will look and feel like a conversation with a trusted friend. Natural Language Processing tries to train AIs to do just that.[176] The AI could carry on a natural conversation about the decision you are facing and respond to your changing mood appropriately. This level of sophistication would

---

172. Stéphane Raffard et al., *Humanoid Robots versus Humans: How is Emotional Valence of Facial Expressions Recognized by Individuals with Schizophrenia? An Exploratory Study*, 176 SCHIZOPHRENIA RES. 506, 506-507 (2016) ("[H]umans have a tendency to attribute human traits to robots, especially when they display human-like physical and behavioral characteristics.") (internal citations omitted); Kerstin S. Haring et al., *FFAB—The Form Function Attribution Bias in Human–Robot Interaction*, 10 IEEE TRANSACTIONS ON COGNITIVE & DEVELOPMENTAL SYS. 843, 843 (2018) (finding that people tend to think robots that look more like people have better AI); Matt McFarland, *Magic Leap's New AI Assistant Looks Alarmingly Human*, CNN BUSINESS (Oct. 12, 2018, 1:47 PM), https://www.cnn.com/2018/10/12/tech/magic-leap-ai-assistant/index.html [https://perma.cc/8QS5-HY6F].

173. MCFARLAND, *supra* note 172; *see also ObEN Gives AI Assistant a Human Face*, MEDIUM (Mar. 7, 2017), https://medium.com/oben/oben-gives-ai-assistant-a-human-face-c0c48caf3eb8 [https://perma.cc/3MQP-BBC9]; *see also* Ye Pan & Anthony Steed, *A Comparison of Avatar-, Video-, and Robot-Mediated Interaction on Users' Trust in Expertise*, 3 FRONTIERS ROBOTICS & AI 1, 2 (2016).

174. *See, e.g.*, EX MACHINA (Film4 Productions 2014) (romantic love); HER (Annapurna Pictures 2013) (same); SOLO: A STAR WARS STORY (Lucasfilm Ltd. 2018) (same); BLADE RUNNER (Warner Bros. Pictures 1982) (same); I, ROBOT (Overbrook Ent. 2004) (companionship); NEXT GEN (Alibaba Pictures 2018) (childhood friendships); A.I. ARTIFICIAL INTELLIGENCE (Amblin Ent. 2001) (parental love).

175. *See* PASQUALE, *supra* note 62, at 71-73 (worrying that if robots replace teachers for children, those children might develop emotional responses and connections to robots, and thereby confuse the important distinction between people and machines).

176. Julia Hirschberg & Christopher D. Manning, *Advances in Natural Language Processing*, 349.6245 SCI. 261, 261-63 (2015). A recent AI can at least begin to have a convincing conversation, but it can become nonsensical in part because the AI cannot yet "remember" what it said to you 10 seconds ago. Cade Metz, *Meet GPT-3. It Has Learned to Code (and Blog and Argue)*, N.Y. TIMES (Nov. 24, 2020), https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html?referringSource=articleShare [https://perma.cc/4ZFV-TSSA].

be ideal.[177] But there are ways to design personalized messages to work better even with less sophisticated interfaces.

Personalized messages should be "layered." The goal here is to create a rough approximation of a natural conversation like you might have when seeking advice from a friend. There are an infinite number of ways to slice up a conversation, so the examples below are not the only possible way to design each layer. But the concept of layering can make personalized messages significantly more effective.

One layer could be what personalized disclosures are currently envisioned to be: simple factual disclosures that highlight information that is relevant to a particular decision. This is where advocates of personalized disclosures have stopped. But drawing on the advice paradigm, we should add more layers.

1. AI Advice and Guinea Pig Advice

Advice would be one of the (perhaps many) layers. Again, we might imagine a simple interface that disclosed information and then allowed users to click, "What should I do?" Like in natural advice settings, this advice could reduce decision fatigue and would not require much literacy or numeracy to understand.

In the medical context, bioethicists have explored the possibility of creating an algorithmic "Patient Preference Predictor" which would help doctors and families make decisions for incapacitated patients.[178] These Personalized Preference Predictors could be adapted to give advice and to give it to competent patients. Such advice might look like: "You should probably get the less invasive surgery because most people like you prefer it, and given your consistent exercise schedule, I think you would value its faster recovery time." Even if the AI could not make the difficult tradeoffs for you, or there was not enough information available to produce a single reliable recommendation, an AI could usefully guide your decision making. Should you prefer living for only three more months or living for six more, but with the constant nausea that can accompany some chemotherapies? An AI might ask a series of questions that are designed to help you value the relevant outcomes. Because people are more honest when answering questions mediated through technology, rather than answering a question

177. *But see* PASQUALE, *supra* note 62, at 8 ("When chatbots fool the unwary into thinking that they are interacting with humans, their programmers act as counterfeiters, falsifying features of actual human existence to increase the status of their machines. When the counterfeiting of money reaches a critical mass, genuine currency loses value. Much the same fate lies in store for human relationships in societies that allow machines to freely mimic the emotions, speech, and appearance of humans.").

178. Annette Rid & David Wendler, *Treatment Decision Making for Incapacitated Patients: Is Development and Use of a Patient Preference Predictor Feasible?*, 39 J. MED. & PHIL. 130, 131 (2014).

directly asked by another person,[179] the AI may do a better job than doctors at facilitating these decisions.

Even unsolicited advice could be a useful layer. Natural advice is often unsolicited. Unsolicited advice is often less effective precisely because the advisee is not in the right state of mind to receive it. She may respond with "mind your own business!" rather than "interesting point!" But this is not always the case, and even if unsolicited advice is always discounted, it still may have some impact. Perhaps when the consumer is particularly tired, or particularly tempted, or otherwise more unlikely to read disclosures, unsolicited advice might appear first.[180] "Don't do it," would be the starting point, followed by factual disclosures explaining the reasons behind the advice.

Although there are objections to unsolicited advice that do not apply to solicited advice, neither is particularly problematic. If a consumer asks for advice, it is hard to imagine an objection to giving it to her so long as the consumer knows who she is asking and is correct in her assumption that they have no conflicts of interest. Unsolicited advice is a bit different. This potentially unwelcome piece of information might have a subtle influence on people. Here, the advice might be acting more like a subconscious nudge. If this is seen as too manipulative, we could always allow people to opt out of unsolicited advice in advance. This may be a sufficient response.

The content of the advice could come from various sources. As this Article has argued, it could come from an AI that has access to vast amounts of data about you. It could also come from what Ariel Porat and Lior Stralhelizitz have called "guinea pigs."[181]

> Under such an approach, a small portion of the population is given a great deal of information and time to make decisions, and then the remaining members of the population are assigned the default terms chosen by the guinea pigs whose observed behavior and characteristics most closely match their own.[182]

Porat and Stralhelizitz suggest that this large-scale dataset could be used to create personalized default rules and personalized disclosures.[183] But Part II revealed the numerous ways that advice is a better paradigm. Luckily, the same infrastructure of guinea pigs can be used to generate defaults, disclosures, or advice. For example, if

---

179. Matheny, *supra* note 5, at 62.

180. Inbal Nahum-Shani et al., *Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support*, 52 ANNALS OF BEHAVIORAL MED. 446, 446-47 (2018) (describing systems that seek to identify moments when patients are particularly in need of intervention and provide those interventions).

181. Porat & Strahilevitz, *supra* note 2, at 1419-21, 1450-53.

182. *Id.* at 1421.

183. *Id.* at 1422.

people "like you" tend to get fifteen-year mortgages instead of thirty-year mortgages, you might receive that advice.

The idea of guinea pig advice is both a partial substitute and a partial complement to AI advice. It is a substitute in that we would no longer be relying on AI to tease out the consequences of a particular action to determine if it is the most welfare-enhancing. We would trust that the guinea pigs are making the most welfare-enhancing choices. But in many areas, the AI may outperform the guinea pigs. AI can already diagnose cancers and predict divorce as accurately or more accurately than experts. It might similarly outperform highly motivated lay people.

Even if guinea pig advice outperforms AI advice in a given area, AI has a role to play. Here, guinea pig advice is a complement to AI advice because we could use AI to determine who is "like you." Porat and Stralhelizitz base this determination on demographics and personality traits.[184] But AI can do much better. The determination that people are "like you" should be context specific rather than stable across persons (who will have stable personality and demographic traits). John and Sally may be similar when it comes to financial decisions, but quite different when it comes to medical decisions. They might have similar discount rates, but very different savings preferences. Using AI to determine who is "like you" can capture these subtleties.

2. Relaying Reasons vs. Explainable AI

Another layer would be reason-giving. Using a simple interface, a recipient of AI advice might be able to click: "Tell me why you think I should do X."[185] This would trigger another layer which could include information about the recipient's past behavior, more information about the consequences of each choice, or often very effective narrative stories that illustrate the relevant dangers.[186] In the context of

184. *Id.* at 1461, 1469.

185. *See* Ryan Browne, *Facebook Has a New Tool That Explains Why You're Seeing Certain Posts on Your News Feed*, CNBC (April 1, 2019, 8:03 AM), https://www.cnbc.com/2019/04/01/facebook-new-tool-explains-why-am-i-seeing-this-post-on-news-feed.html [https://perma.cc/8ZVC-7RSB].

186. John B.F. de Wit et al., *What Works Best: Objective Statistics or a Personal Testimonial? An Assessment of the Persuasive Effects of Different Types of Message Evidence on Risk Perception*, 27 HEALTH PSYCHOL. 110, 110, 113 (2008) (finding that narratives work better at increasing risk perception than statistics); Victoria A. Shaffer & Brian J. Zikmund-Fisher, *All Stories Are Not Alike: A Purpose-, Content-, and Valence-Based Taxonomy of Patient Narratives in Decision Aids*, 33 MED. DECISION MAKING 4, 4 (2013) (noting that the use of narratives to educate patients is controversial precisely because they are thought to be too influential compared to statistics); Hilary L. Bekker et al., *Do Personal Stories Make Patient Decision Aids More Effective? A Critical Review of Theory and Evidence*, 13 BMC MED. INFORMATICS & DECISION MAKING S9, 7 (2013) (finding that "[p]ackaging information as a story affects people's use of both system 1 (intuitive-experiential) and system 2 (deliberative-analytic) information-processing strategies"). Of course, there are numerous other options. For example, messages could also include graphic images, or be embedded in video games.

personalized disclosures, Oren Bar-Gill has proposed disclosing facts about the consumer's own past behavior.[187] The consumer herself may have honestly forgotten that she incurred late-payment penalties twice last year, or the consumer may have blocked out that information through motivated reasoning.[188] Bar-Gill envisioned these as stand-alone disclosures,[189] but they could be useful in a reason-giving layer following advice as well. Other reasons would help too: "You should avoid the credit card cash advance, *because* I think you are considering it in order to pay off your electric bill, and you probably qualify for a bill forgiveness program" or " . . . because there's too big a chance that something will come up beyond your control and you'll have to borrow even more later."[190]

Reason-giving is importantly distinct from calls for AI to offer greater levels of "explainability."[191] The key difference is that explainability normally targets different audiences and has different goals than reason-giving.

Numerous scholars have sought ways to open up the black box of AI to allow for humans to assess its output.[192] Three ideas dominate the landscape of this explainable AI (XAI) movement.[193] First, creators of AI could restrict the number of input variables or restrict the number of branches in its decision tree.[194] This would simplify the underlying algorithms and make the AI's outputs more discernable to

187. Oren Bar-Gill, *Consumer Transactions*, *in* THE OXFORD HANDBOOK OF BEHAVIORAL ECONOMICS AND THE LAW 480-81 (2014).

188. *See supra* notes 90-97 and accompanying text.

189. *Id.*

190. Of course, revealing such information may promote reactance and defensiveness. Here, packaging the information in an emotionally cognizant way could be useful. *See infra* Section III.C.

191. Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1085 (2018); *see* DEF. ADVANCED RES. PROJECTS AGENCY, BROAD AGENCY ANNOUNCEMENT: EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) (2016).

192. *See, e.g.*, Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 6 (2014); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1280 (2008); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1346-53 (2018); Susan C. Morse, *When Robots Make Legal Mistakes*, 72 OKLA. L. REV. 213, 230 (2019). The importance and type of explainability that might be desirable can affect the design of the AI. Many machine learning techniques today take advantage of neural networks. Neural networks can solve immensely complicated problems and are the root of successes in training AI to compete in Chess and Go. But these neural networks are so complex that people cannot understand why they do what they do. Even if a neural network could explain itself, it would hardly value parsimony and clarity. Its explanation may be 647 pages long and include thousands of interrelated correlations. Some alternatives to these black box neural networks exist that offer greater potential for reason-giving, and progress is being made to increase the explicability of even deep neural networks. David Gunning & David W. Aha, *DARPA's Explainable Artificial Intelligence Program*, 40 AI MAG. 44, 45, 47 (2019).

193. Selbst & Barocas, *supra* note 191, at 1110.

194. *Id.* at 1110, 1111.

people. Second, creators of AI could embrace a dual-model approach, where the AI would generate two different models.[195] A complex model would control its actual output.[196] But a simplified version would be used to explain that outcome.[197] This simplified model would get the gist but not the details correct.[198] Third, people could have access to simulators that allow them to alter parameters to see how the underlying algorithm works.[199] For example, Danielle Citron and Frank Pasquale argue for "interactive modeling."[200] Under this idea, an AI credit score calculator might allow people to input hypothetical choices and see how those choices affect their credit score.[201]

Explainability serves multiple goals for multiple constituencies. First, people who are affected by an AI's output might have dignitary interest in understanding them.[202] That is, they might have a dignitary interest in understanding why an algorithm assigned them a low credit score, denied them government benefits, or declined to pass their resume on to a human-decider. Second, these people have more tangible interests in these determinations as well.[203] Credit scores, government benefits, and employability scores could all be vital to opening or closing off important opportunities. Accordingly, people have an interest in understanding how they can open up these opportunities.[204] Third, governments have an interest in ensuring that algorithms do not discriminate on the basis of race, religion, sex, or other features that it deems illegitimate to making the determination at hand.[205] Fourth, the entities that create or use AI (be they governments, companies, scientists, etc.) have an interest in making sure that it is working properly and not, for example, confusing

195. *Id.* at 1113.

196. *Id.*

197. *Id.*

198. *Id.* at 1110, 1113.

199. *Id.* at 1110, 1115-16.

200. Citron & Pasquale, *supra* note 192, at 28-29.

201. *Id.*

202. *Id.* at 27.

203. *Id.*

204. *Id.*

205. *See* Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459, 460 (2019) (noting the difficulty of policing the inappropriate use of race or other demographic characteristics when humans cannot understand the underlying algorithm).

correlation with causation.[206] Explainability serves these and many more goals.[207]

Reason-giving in the advice context serves a different goal. The primary goal is simply to increase the probability that the person will act in a welfare-enhancing way.[208] Reason-giving assists this goal by helping people verify that the AI understands their situation and goals, and by simply making people feel more comfortable with AI advice. Some people will be suspicious that automated advice will be ill-informed or just bad. Reason-giving is one way that AI advice can overcome these barriers and build trust. Dignitary interests seem significantly less important, as do concerns about Kafka-esque algorithms controlling one's life. One's dignity and power are reaffirmed by the very structure of advice: you are still in control over whether you take the advice and allow it to affect your life or opportunities.[209]

These different goals affect the choice of how to promote explainability. Consider again the analogy to natural advice. How do advisors explain their advice? They do not constrain the inputs they use to ensure that they can articulate the justifications behind their advice. They may offer something like a simulator, in that they can respond to questions from the advisee that might include counterfactuals. But impersonal forms of interactive modeling carry risks. Motivated reasoning might well cause people to fiddle with the inputs until they get the result that they want. In the context of natural advice, the friend-advisor can combat the allure of this motivated reasoning. But interactive modeling on a smart phone would not have this counterweight to motivated reasoning.

The dual-model approach provides a good start. When the AI provides reasons, it can simplify. The reasons might, for example, only refer to the three most important factors in determining the advice.

---

206. Lemley & Casey, *supra* note 6, at 1336-37 (noting an early machine learning conclusion that people with a history of asthma are more likely to survive pneumonia, and the error involved in that conclusion: not accounting for the fact that asthma is such a large risk factor for death from pneumonia that asthmatics are automatically put into intensive care if they have signs of pneumonia); Citron, *supra* note 192, at 1305 (suggesting that AI keep "audit trails that record the facts and rules supporting their decisions").

207. Explainability may be important to existing legal rules. For example, a person may need to know why an AI flagged their computer as possibly containing child pornography in order to assess whether a warrant issued in reliance on that determination violated his Fourth Amendment rights. Explainability may also be related to perceived legitimacy.

208. See Section II.B. for slight variations on this theme. See Section IV.A. for other potential goals of reason-giving, for example to build cognitive capacity.

209. *See* Citron & Pasquale, *supra* note 192, at 25 (arguing that we should worry most when people "can do little or nothing to protect themselves" from the negative effects of an AI's determination). The government's interests in ensuring accuracy and avoiding discrimination are assuredly still relevant, but they would require more than just casual reason-giving that average lay people can understand and appreciate.

Under this constraint, a response to "why did you give me this advice" might look something like:

> Fifty-two variables each increased the odds of one of fourteen potential negative events. The three most important are your predicted income over the next month, which increased the odds of not being able to make the required balloon payment by .37, your repayment history, which increased the odds of a late payment by .14, and the low conscientiousness score on your personality-trait inventory, which increased the odds of accidentally missing a payment by .08.[210]

This may be useful to scientists who are designing the AI or assessing its accuracy, but it seems like a poor choice for convincing someone that the advice is sound. It could trigger overload effects, defensiveness, or just seem too mechanical to be trustworthy.[211] A more fruitful version might use significantly vaguer language to highlight some aspects of the underlying algorithm and downplay others:

> You won't make enough money this month to afford that balloon payment, and even though you are very conscientious, you missed a few monthly payments last year because of forces outside of your control.

Sometimes, simple (and supportive) is best.[212]

3. Precommitments

Another layer could offer precommitment devices. A recent field experiment with almost 50,000 patients attempted to improve the rate that they took their medications.[213] It compared the effects of sending basic reminders, asking people to predict their medication compliance, and asking them to commit to themselves that they would achieve any self-determined level of compliance.[214] Asking people to precommit was the only treatment that improved compliance both during the reminder phase and after those reminders ended.[215] Dozens of other studies confirm the power of prompting people to make informal commitments or even just to make informal plans to do something. Even just asking people about their future intentions affects behavior.

---

210. *See* Selbst & Barocas, *supra* note 191, at 1103.

211. *See id.* ("If the creditor tried to explain [credit scoring] rules simply, it would leave information out, but if the creditor were to explain in complete detail, it would likely overwhelm a credit applicant.").

212. Kartik Hosanagar & Vivian Jair, *We Need Transparency in Algorithms, But Too Much Can Backfire*, HARV. BUS. REV. (July 23, 2018), https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire [https://perma.cc/9L5Y-UXAK] (noting in a natural experiment that people "trusted the algorithm more when they got more of an explanation of how it worked" but only up to a point; too much information about the algorithm decreased trust).

213. Hengchen Dai et al., *The Effect of Interactive Reminders on Medication Adherence: A Randomized Trial*, 103 PREVENTIVE MED. 98, 98-99 (2017).

214. *Id.* at 99-101

215. *Id.* at 100-01.

Asking people about whether they intend to donate blood, get a flu shot, vote, recycle, exercise, buy a computer, or donate to charity increases their probability of doing so later.[216] Similarly, asking people about whether they will cheat on an exam reduces the likelihood that they will.[217] A layered message could easily incorporate some or all of these mild precommitments.

Stronger precommitments could be included as well. The mild precommitments above do not rely on the threat of any external costs being imposed if you fail to stay the course. It would be easy to add such costs. The website Stickk.com offers people the opportunity to create monetary stakes in their own future behavior. You can, for example, give $500 to the website, and you will only get it back if you accomplish your goal.[218] AI could offer similar precommitments, and in fact could offer even more effective versions. Stickk users often ask a third person to sign up as a referee because Stickk itself has no way to check whether users accomplish their goals.[219] But AI will often be able to obtain this information with big data. It will know whether you signed up for a retirement account, and how much you are contributing. Your smart watch will tell it whether your BMI has decreased sufficiently. Accordingly, in many cases, it would be much easier and faster to create these monetary precommitments through a system like AI advice.

An AI could also use social costs to enforce precommitments. One could precommit to an automated email or text going to friends or family upon some triggering event. Perhaps every time you go to a casino, your spouse gets a text.[220] Or every time you buy ice cream,

---

216. Eric R. Spangenberg et al., *A Meta-Analytic Synthesis of the Question–Behavior Effect*, 26 J. CONSUMER PSYCHOL. 441, 442-43 (2016); Gaston Godin et al., *Asking Questions Changes Behavior: Mere Measurement Effects on Frequency of Blood Donation*, 27 HEALTH PSYCHOL. 179, 180-82 (2008). Of course, there are caveats. The effect is smaller when the behavior is more difficult. The effect is larger for more socially desirable behavior. Chantelle Wood et al., *The Impact of Asking Intention or Self-Prediction Questions on Subsequent Behavior: A Meta-Analysis*, 20 PERSONALITY & SOC. PSYCHOL. REV. 245, 245, 260-61 (2016).

217. Eric Spangenberg & Carl Obermiller, *To Cheat or Not to Cheat: Reducing Cheating by Requesting Self-Prophecy*, 6 MARKETING EDUC. REV. 95, 95 (1996).

218. STICKK, https://www.stickk.com [https://perma.cc/M2SS-E9HM] (last visited May 12, 2021).

219. *Id.*

220. Relatedly, one alcohol recovery program gave subjects a smartphone that issued them a warning when its GPS indicated that they were near a bar they used to go to. *See generally* David H. Gustafson et al., *A Smartphone Application to Support Recovery from Alcoholism: A Randomized Clinical Trial*, 71 JAMA PSYCHIATRY 566 (2014). It also included a "panic" button, which would contact designated others. *Id.* at 568. More complex algorithms can also be used to predict when someone is in need of, and will be receptive to, a particular intervention. Benedict Carey, *Can an Algorithm Prevent Suicide?*, N.Y. TIMES (Nov. 23, 2020), https://www.nytimes.com/2020/11/23/health/artificial-intelligence-veterans-suicide.html?referringSource=articleShare [https://perma.cc/PTK4-NFPE] (describing Department of Veteran's Affairs AI which identifies veterans who are at risk of suicide); PASQUALE, *supra* note 62, at 73 ("Facebook, for instance, can algorithmically detect postings by teens that indicate suicidality and refer them to specialized help.").

your personal trainer gets an email. Your friends may get a text if you get a raise but do not increase your retirement contributions proportionately. These are just some of the ways that an AI could leverage social forces.

Finally, in a truly connected world there may be many more ways to precommit. In response to the problems of gambling addiction, several states have offered self-registries that bar people from entering casinos.[221] Extrapolating from this example, one could imagine a world in which people could sign up for a fringe credit freeze which would bar payday lenders and pawn shops from loaning them money for a certain time. Or, with a few taps on a smart watch, they could temporarily add themselves to a federal registry that bars them from buying a gun, which they might want to do if they are susceptible to suicidal thoughts.[222] Finally, people might be able to put certain items on a "no buy" list so that their credit card would decline any transaction that included them. Quick and easy ways to enter these and other precommitments expands choice and allows people with bounded willpower to have more control over their lives.

4. Nudging

Layers could also include nudging. Here, I am not using Thaler and Sunstein's capacious definition of nudge, under which every layer would be a nudge.[223] Rather, I am referring to a less well-defined subset that relies more on subtle influence than rational persuasion. There is no single answer to where in the series of layers a nudge like this might occur. For some decisions or for some people, it might occur as part of an early layer. One could imagine a disclosure outlining the dangers of an auto-title loan, followed by a default rule of sorts. If you actively click "proceed despite the warning" you can get the loan. But if you do not click anything, the loan could not be processed. The wording of this could moderate the strength of the default rule. It is easier to click "proceed anyway" than to click "do the thing that most people will regret and will probably lead to financial ruin." Rental car companies know this. That is why their employees might ask you if you are sure you want to "remain unprotected" or are "willing to take the risk" before declining their expensive and often unnecessary insurance coverage.[224] Rental car companies also try to use social norms and defaults to influence choice by saying that the largely

221. Ian Ayres & Fredrick E. Vars, *Libertarian Gun Control*, 167 U. PA. L. REV. 921, 944 (2019).

222. *Id.* at 921.

223. THALER & SUNSTEIN, *supra* note 11, at 6.

224. Tom Baker & Peter Siegelman, *"You Want Insurance with That?" Using Behavioral Economics to Protect Consumers from Add-on Insurance Products*, 20 CONN. INS. L.J. 1, 2 (2013).

useless and expensive damage waiver is the "[m]ost popular" choice.[225] Layered disclosures could similarly use social norms. Big data could discover what those norms actually are for the general population, for people like the individual in question, or even for the individual's actual Facebook friends. When revealing those norms increases the likelihood of welfare-enhancing decisions, they could be included in one of the layers.[226] For example, a layered disclosure might ask "Do you want to know what your friends do in this situation?"[227] and then reveal that they do not generally purchase the collision damage waiver.[228]

For purposes of this Article, one particularly noteworthy nudge entails reverse psychology and purposefully giving bad advice. Some people are likely to see advice as threatening their independence and freedom.[229] This can lead them to do the opposite of what the advice recommends. We are all familiar with this tendency. The strategy of reverse psychology tries to take advantage of it.[230] AI advice could attempt to predict who is likely to push back against advice and offer bad advice in the hopes that the target will do the opposite. This strategy has many of the same objectionable features as classic nudges, such as the pall of manipulation. But if those are acceptable, then reverse psychology advice should at least not be dismissed out of hand as a potentially useful new type of nudge to add to a policymaker's toolkit.[231]

---

225. DOLLAR, https://www.dollar.com/ [https://perma.cc/UG67-WTXF] (last visited May 12, 2021); *see also* THRIFTY, https://www.thrifty.com/ [https://perma.cc/99V8-R3JB] (last visited May 12, 2021) (including loss damage waivers in its "Basic" package rather than in its other coverage packages).

226. Here, I set aside normative questions about whether to hide or reveal non-welfare-enhancing norms, or whether to lie about those norms. *See* Cass Sunstein, *Sober Lemmings*, NEW REPUBLIC 34 (April 14, 2003) (engaging with some of those issues).

227. Here, I set aside privacy concerns.

228. Nudging could also occur in other layers and use less classic nudging techniques. For example, simply altering the wording of advice could make it easier or harder to ignore. "I'd avoid the loan, but reasonable people could disagree" is different than "Only an idiot would get this loan." Of course, numerous other nudges are possible as well.

229. Gavan J. Fitzsimons & Donald R. Lehmann, *Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses*, 23 MARKETING SCI. 82, 82-83 (2004); Anna Goldfarb, *How to Give People Advice They'll be Delighted to Take*, N.Y. TIMES (Oct. 21, 2019), https://www.nytimes.com/2019/10/21/smarter-living/how-to-give-better-advice.html?smid=nytcore-ios-share [https://perma.cc/A949-E2RV] ("Sometimes when people don't take advice, they're rejecting the idea of being controlled by the advice-giver . . . .").

230. Geoff MacDonald et al., *Do People Use Reverse Psychology? An Exploration of Strategic Self-Anticonformity*, 6 SOC. INFLUENCE 1, 11 (2011).

231. Of course, policymakers would also have to consider numerous other factors, such as whether people will find out that the AI uses reverse psychology and whether this will undermine its perceived credibility.

C. *Emotionally Cognizant Layers*

Each step of a layered message could be attentive to the emotional context surrounding the relevant decision.[232] As explained above, friends are likely to know when some particular piece of information, line of argument, or offered advice is threatening, and can respond accordingly. Simply paying attention to the emotional context can mitigate motivated reasoning, reduce overoptimism, promote self-efficacy, and increase the likelihood of altering behavior for the better.[233] AI advice can, and should, do the same. In fact, AI might eventually perform better than friends, both because it has access to more data and because it can conduct numerous experiments about which techniques work best.

Threatening information can be packaged in ways that make it easier to digest. Affirmations and encouragement may help here.[234] The AI could point out that time when you donated to a local charity or could comb through your Facebook posts to find one where you were a good friend to someone in need. Reminding people of their previous acts of kindness helps affirm their sense of self-worth.[235] As discussed above, this can reduce resistance to threatening information.[236]

Reason-giving can also be designed with emotional sensitivity. Even if a person has a long history of not paying their credit card bills on time, it might not be a good idea to throw that data back in their face.[237] A better strategy might be to highlight risks beyond their control that might prevent them from paying off a loan on time. Highlighting the ways that unexpected costs can appear is less threatening than highlighting the consumer's own past lack of self-control. It has the added benefit of reducing overoptimism, which tends to be much stronger when people have a sense of control over the relevant outcome.[238]

Advice might be more effective, and consumers may be more likely to precommit, if those layers include various encouragements. Again,

---

232. Schaller & Malhotra, *supra* note 84, at 691-92 (criticizing medical disclosures for not being sufficiently attentive to emotionally charged context of the relevant decisions).

233. *See supra* Sections II.C., II.D. (motivated reasoning, self-efficacy, behavioral change); *see infra* note 233 and accompanying text (over-optimism).

234. *See* Karáth, *supra* note 32("[U]sing almost 190,000 questions and 770,000 answers from [one relationship advice website]. . . . [t]hey came up with a generic structure for answers that includes a sentence showing sympathy, a suggested solution to the problem, an additional comment and a note of encouragement.").

235. *See* Mark B. Reed & Lisa G. Aspinwall, *Self-Affirmation Reduces Biased Processing of Health-Risk Information*, 22 MOTIVATION & EMOTION 99, 107, 122-23 (1998).

236. *See supra* Section II.D.

237. Simine Vazire & Erika N. Carlson, *Others Sometimes Know Us Better Than We Know Ourselves*, 20 CURRENT DIRECTIONS PSYCHOL. SCI. 104, 107 (2011) ("Direct, honest feedback might be very useful, but it is rare, and probably for good reason.").

238. David Dunning et al., *Flawed Self-Assessment: Implications for Health, Education, and the Workplace*, 5 PSYCHOL. SCI. PUB. INT. 69, 80 (2004).

an AI may have access to information that can make these encouragements more effective. "You can do it!" may help, but the following might be more effective: "You can do it! Remember last week when you hit your screen time goals? That was hard and you did it. You can do this too." Recall also the phenomenon of self-talk, which boosts athletic performance and helps people persevere in the face of challenges.[239] Although AI cannot talk for you, an emotionally cognizant layer of AI advice might encourage this form of self-talk.[240]

Even just communicating sympathy can make messages more potent. In one recent study, subjects rated the helpfulness of various statements that a friend might make in the course of helping them choose a job.[241] Unsurprisingly, subjects rated recommendations ("I think you should pick job C") as helpful.[242] Perhaps less obviously, statements of generic emotional support ("I understand that this is a really tough decision for you; choosing a job is a very stressful decision") were just as helpful.[243] As the developer of one consumer call chatbot has already realized, expressing sympathy can also help disarm anger and lead to more productive conversation.[244]

Many more psychological customizations are possible. Consider framing. For some people, or in some situations, it might be more effective to focus on the negative rather than the positive. Friends are likely to know when cajoling is in order, and when teasing would work better. AI could learn similar nuance. For example, some people may respond more to losses: "You will be $50 deeper in debt if you get that loan." Other people may respond better to gains: "You will have $50

239. *See supra* note 125.

240. Of course, the optimal level of emotional engagement may differ depending on whether the advisor is a friend or an AI. A friend who points out the way that you were kind last week shows that they are attentive and caring. AI that does this may show that it is creepy and invasive. *See* Alexander Bleier & Maik Eisenbeiss, *The Importance of Trust for Personalized Online Advertising*, 91 J. RETAILING 390, 395, 402-03 (2015) (finding that consumers embrace ads with personalized content when they originate from a company they trust but respond negatively to similarly personalized ads from companies that they do not trust). A friend who says, "I know you can do it!" is showing their support and reinforcing how well they know you. When an AI says this, it might feel empty. This suggests that there are likely gains from incorporating more emotional context, but that more information is needed to assess the optimal level of emotional sensitivity. Eventually, AI may be able to gather this information with enough trial and error and settle on levels of emotional support that are somewhere between none and the level that a friend would provide.

241. *See* Reeshad S. Dalal & Silvia Bonaccio, *What Types of Advice Do Decision-Makers Prefer?*, 112 ORG. BEH. & HUM. DECISION PROCESSES 11, 14 (2010).

242. *Id.* at 15, 17.

243. *Id.*; *see* Allison C. Shields, *Simple Steps: Emotional Intelligence versus Artificial Intelligence*, ABA (July/Aug 2019), https://www.americanbar.org/groups/law_practice/publications/law_practice_magazine/2019/JA2019/JA2019SimpleSteps/ [https://perma.cc/25WE-TL6S] ("An injured client or a grieving family chooses a legal service provider based on how that provider makes them feel, not just whether they think the provider can get them the best legal or financial outcome for their case. As a result AI can aid lawyers by recommending legal strategies, but machines likely won't completely replace lawyers.").

244. *See* Friedman, *supra* note 170.

extra at the end of next week if you use the other loan instead." Under the assumptions that started the discussions of personalized disclosures, a smart AI with access to immense troves of personal data, these personalizations are certainly possible. The AI could even run its own experiments on which frames work best and for whom.[245]

D. Summary

Looking at natural advice reveals a number of insights. Many of which can be incorporated into AI advice. In addition to offering accurate advice, an AI can offer disclosures and other reasons that shaped the content of the advice. It can offer precommitment devices that mimic those seen in natural advice, and it can design each layer in ways that promote self-efficacy and minimize motivated reasoning.

IV. OBJECTIONS AND EXTENSIONS

This section will set aside a series of standard concerns that others have already discussed at length. For example, we might worry about something akin to agency capture if the government began to develop and implement a system of AI advice.[246] We might worry about discriminatory algorithms,[247] or algorithms that exacerbate the digital divide,[248] or how to regulate private-sector AIs.[249] We might worry

245. "Best" could be defined as taking the AI's advice, or as having less debt six months later, etc. Human input can guide the AI's definition of "best."

246. *See generally* PREVENTING REGULATORY CAPTURE: SPECIAL INTEREST INFLUENCE AND HOW TO LIMIT IT (Daniel Carpenter & David A. Moss, eds., 2014). Some of the harms of delegating power to an agency are not present for AI advice. The technological optimism assumed by this Article alleviates concerns that that policymakers will be as error-prone as the consumers, or will operate under the wrong incentives, and hence, should not attempt to influence them. *See* Mario J. Rizzo & Douglas Glen Whitman, *Little Brother is Watching You: New Paternalism on the Slippery Slopes*, 51 ARIZ. L. REV. 685, 686 (2009); Edward L. Glaeser, *Paternalism and Psychology*, 73 U. CHI. L. REV. 133, 142, 144-49 (2006); Aneil Kovvali, *Who Are You Calling Irrational?*, 110 NW. L. REV. 707, 708 (2016).

247. Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389, 1394-1406 (2019). Some of the harms of algorithms are not present in AI advice. YouTube and Facebook have algorithms that arguably contribute to political polarization and conspiracy theories. They seek to maximize views, and people love hearing things they already agree with. If AI advice sought to maximize the uptake of its advice, or maximize the immediate satisfaction with the advice given, then it might start telling people only what they want to hear. But AI advice need not be this short-sighted. Its goal is to offer good advice, not agreeable advice. The accuracy of AI advice would be continually reassessed in the same way it was created: by examining correlations between actions and verifiable outcomes like financial distress, health problems, etc. Of course, one of those outcomes could include the narrow satisfaction that a particular piece of advice triggers, but even if so, it would be only one of many.

248. Peter K. Yu, *The Algorithmic Divide and Equality in the Age of Artificial Intelligence*, 72 FLA. L. REV. 331, 334 (2020).

249. Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. 633, 695 (2020) ("AI development is an especially stark example of how private coding choices are governance choices, embedded within products that affect both norms and physical wellbeing. . . . [T]his Article suggests that policymakers should focus on how to filter public

about "black box" decisions that lack transparency,[250] and we might seek to offer "technological due process" for advice recipients.[251] Instead of rehashing these debates, this Part will discuss a concern that may at first seem especially poignant for AI advice: Will we come to rely on it so much that our cognitive abilities to critically examine our world will atrophy? The answer is no. It then argues that even if policymakers severely constrain or reject AI advice, many of the lessons generated by this Article can, and should, shape the future of mandatory disclosure and nudges.

A. Will AI Advice Make Us Dumb?

Perhaps AI Advice will be so good that it will erode people's capacity to think through complex problems on their own. In short, people might worry that it will make us dumb.

These worries are not unique to AI advice. Socrates thought that literacy and writing would erode people's ability to remember.[252] Others have wondered whether search engines would erode our ability to remember facts.[253] Still others have wondered whether consumer protections would impair cognitive ability by delinking mistakes from consequences.[254] We might also worry that mandatory disclosures would erode our capacity to discover or sift through information. Evidence for these negative consequences is difficult to find,[255] and any

---

input and instill public values through . . . markets and norms, rather than attempting direct control of algorithmic technologies like AI through the law."); Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 122 (2017) (advocating for "a new federal agency whose goal is to ensure that algorithms are safe and effective").

250. Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 5 (2019) ("Can algorithmic governance be squared with legal demands for transparency?").

251. Citron, *supra* note 192, at 1300-01.

252. Maria Konnikova, *On Writing, Memory, and Forgetting: Socrates and Hemingway Take on Zeigarnik*, SCI. AM. BLOG (Apr. 30, 2012), https://blogs.scientificamerican.com/literally-psyched/on-writing-memory-and-forgetting-socrates-and-hemingway-take-on-zeigarnik/ [https://perma.cc/M9P5-TYEW] (quoting The Phaedrus) ("[T]hey will be hearers of many things and will have learned nothing; they will appear to be omniscient and will generally know nothing; they will be tiresome company, having the show of wisdom without the reality.").

253. *See* NICHOLAS CARR, THE SHALLOWS: HOW THE INTERNET IS CHANGING THE WAY WE THINK, READ AND REMEMBER 120 (2010).

254. *See* Jonathan Klick & Gregory Mitchell, *Government Regulation of Irrationality: Moral and Cognitive Hazards*, 90 MINN. L. REV. 1620, 1633, 1636-37 (2006); Douglas Glen Whitman & Mario J. Rizzo, *Paternalist Slopes*, 2 NYU J.L. & LIBERTY 411, 431 (2007) (worrying about "paternalist slopes" which occur when the "failure to learn self-management techniques can lead to more errors of judgment, which then are used to justify further interventions"); Stephanie Plamondon Bair, *Malleable Rationality*, 79 OHIO ST. L.J. 17, 20 (2018) (noting that "individual decision-making *capacity*—is malleable" and depends on the law-shaped environment in which that capacity develops and is exercised).

255. *See* Jeffrey J. Rachlinski, *How I Learned to Stop Worrying and Love Nudges*, 95 TEX. L. REV. 1061, 1072 (2017) (noting that "[h]ard evidence that would provide direct support of [Klick and Mitchell's cognitive hazard] thesis is hard to come by . . . ."); *but see*

deficits would have to be weighed against potentially offsetting benefits.[256] Further, compared to these other areas, AI advice is perhaps less likely to impair our capacities. As described above, it can include reason-giving. Assessing the strength of these reasons promotes, rather than erodes, reasoning skills.[257] Nonetheless, some readers may still worry about eroded capacities.[258]

It is unlikely that AI advice will be so ubiquitous that it will prevent us from exercising and developing the relevant cognitive capacities. Spaces will remain where AI advice does not operate. These spaces may be sufficient to prevent the loss of our cognitive capital. If there is a nonlinear relationship between opportunities to exercise judgment and building cognitive capital,[259] then we may simply not need all of the opportunities to screw up that life provides. The lessons that one learns in those unprotected spaces may well generalize. As long as some decision domains still require cognitive effort, we may well avoid the cognitive hazard that some worry about.

Various normative disagreements will create one set of spaces for individuals to exercise and build their cognitive capacities. Suppose the AI predicted that you could successfully rob a liquor store, as long as you killed the clerk. Should it recommend doing so? Clearly no. This suggests the need for constraints on the scope of AI recommendations. Constraining it so that it cannot recommend criminal activities is perhaps an obvious place to start. But other areas may create more disagreement. Policymakers might have to make tough choices about

---

John Danaher, *Toward An Ethics of AI Assistants: An Initial Framework*, 31 PHIL. & TECH. 629, 636 (2018) (describing one unpublished study and one published one that found that assistance undermined learning).

256. Danaher, *supra* note 255, at 636 ("But degeneration of performance on certain tasks is not enough, in and of itself, to show that there is a problem. It could well be tasks in which our performance degenerates are not that important in the first place, and freeing us from their performance might be a net benefit.").

257. *See* Sandro Ambuehl, et al., *Peer Advice on Financial Decisions: A Case of the Blind Leading the Blind?* 3 (Nat'l Bureau of Econ. Res., Working Paper No. 25034, 2018) ("[P]eer-to-peer communication does not improve the quality of decision making merely because subjects can identify others who are better informed, without comprehending how the other person arrived at their decision. Instead, communication appears to provide our subjects with generalizable decision skills that are applicable beyond the specific problems they discuss."). These benefits would only occur if people accessed the reason-giving layer. But as natural language processing evolves, the AI may be able to communicate those reasons fairly quickly and automatically.

258. A related concern is that, if AI advice becomes truly excellent, people may place too much trust in it. They might, then, be insufficiently skeptical of, for example, *corporate* AI advice, which might steer them toward decisions that benefit the corporation. But trust in AI advice is likely to be fragile. Although the hype of AI tends to make people trust AI more than humans, they judge AIs more harshly when they fall short of their inflated expectations. Lyn M. Van Swol et al., *Advice Recipients: The Psychology of Advice Utilization*, *in* THE OXFORD HANDBOOK OF ADVICE 21, 33 (Erina L. MacGeorge & Lyn M. Van Swol eds., 2018); Matheny, *supra* note 5, at 90-91.

259. For a discussion of "domain-general" vs. "domain-specific" learning, see Klick & Mitchell, *supra* note 254, at 1631.

how to handle externalities,[260] or constrain the AI to offering advice in situations where externalities are less salient. More broadly, the AI could be constrained such that it could not offer advice for choices where determining the "better" choice requires contested normative judgments.[261] Should you get an abortion? Should you go to church? Should you give to charity? Whom should you marry? Should you lie to your spouse in a way that arguably makes them better off? Should you adopt a free-range philosophy of parenting? Who should you vote for?[262] Even if the AI could accurately determine the hedonic impacts of each of these actions for you,[263] or its monetary impacts, we might worry that other concerns are too salient to ignore, and too contested to resolve.

Just as AI advice might not offer guidance on some of life's biggest questions, it may also have little to say about many of life's smaller questions. Should I walk or bike to the park? Should I try a new type of sushi? Again, these spaces give people numerous opportunities to act, feel the consequences, learn, and grow their decisional capacities. If cognitive hazards are still concerning, policymakers could further constrain the scope of AI advice. The AI might be precluded from offering advice for a set of low-consequence decisions where the relevant lessons would generalize to other contexts. People could then make plenty of mistakes and learn from them in ways that would grow their cognitive capital. A related solution would be to provide AI advice only for important decisions that are rarely made. For example, people do not get many opportunities to learn about mortgages or retirement savings. Accordingly, reducing the learning that occurs in these spaces will not greatly affect one's overall cognitive capital and could help many people amass a great deal more wealth. Even if cognitive

260. For example, someone will have to decide how self-driving cars should weigh the lives of drivers, passengers, and pedestrians. Edmond Awad et al., *The Moral Machine Experiment*, 563 NATURE 59, 59 (2018).

261. John Beshears et al., *How Are Preferences Revealed?*, 92 J. PUB. ECON. 1787, 1793 (2008) ("Like doctors, the government . . . [is] in a good position to advise autonomous agents . . . . Governments could play a constructive advisory role if [ ] their advice is only given in circumstances when the many different measures of normative preferences discussed above tend to coincide . . . . By contrast, in cases with ambiguous or contradictory measures of normative preference, we side with Hayek and Friedman—government should withdraw."). Some of these issues could be mitigated through "interactive modeling" which would allow users to define their own normative commitments explicitly. *See supra* note 200 and accompanying text.

262. This also implicates governmental conflicts of interest. We might not want a particular administration to be able to influence AI advice in ways that make it more likely to advise citizens to keep that administration in power.

263. This should not be read to imply that I endorse a hedonic account of welfare. It is one possible account, and it is one that might be useful in situations where people do not have well-defined preferences.

hazards become more than just a speculative concern, these or other constraints provide ways to avoid it.[264]

B. *Will AI Advice Turn us into Automatons?*

The future of AI and big data could undermine our capacities for critically examining *ourselves*, rather than just undermining our ability to solve complex problems. AI advice, if it is centralized, excellent, and embraced by many, could channel people toward certain products, experiences, and choices. This could reduce the amount of experimentation that people engage in, and this in turn might reduce their capacities to critically examine their own choices, preferences, goals, projects, and identities.[265] This worry is part of a larger set of worries about big data. The constant surveillance of a big data world may chill experimentation and cause people to adhere to social norms more fully.[266] As discussed above, this Article takes a pessimistic view of future privacy protections. It assumes a world of surveillance where our lives are subject to datafication. Given this backdrop, AI advice would probably not add significantly to the problem. But it is at least possible that AI advice will exacerbate the problematic features of this big data world by more explicitly channeling people into well-trodden pathways.[267]

If these worries are well-founded, then one solution is again to constrain the scope of AI advice (or more broadly, constrain the scope of big data surveillance). It is not clear how much space people need to

264. The AI itself could also run experiments to see whether cognitive hazards are significant, and if so, design solutions. To the extent that policymakers place a value on cognitive capacities above and beyond their relationship with individual outcomes (perhaps because they place intrinsic value on cognitive capacities or think that those capacities are important for a functioning democracy), they could constrain the AI to give special weight to developing the cognitive capacities of its users. Accordingly, the AI might offer more boosts. Ralph Hertwig & Till Grüne-Yanoff, *Nudging and Boosting: Steering or Empowering Good Decisions*, 12 PERS. PSYCHOL. SCI. 973, 977, 981 (2017) (defining "boosts" as interventions designed to improve decision-making capacities, sometimes in ways that will generalize to other decision areas). The reason-giving features of natural advice create one such boost. Law students might also imagine an annoyingly Socratic AI, who never tells you the answer but rather tries to get you to come to that answer yourself using opaquely leading questions.

265. Danaher, *supra* note 255, at 643 (critiquing the argument that "AI [recommendations] would gradually 'nudge' you into a set of preferences and beliefs about the world that are not of your own making").

266. Julie E. Cohen, *Power/Play: Discussion of Configuring the Networked Self*, 6 JERUSALEM REV. L. STUD. 137, 143 (2012) (arguing that the autonomous liberal self only emerges and can only be maintained if people have sufficient space to play and experiment, and that because these spaces are eliminated by pervasive surveillance and the perfect enforcement of rules, autonomy requires "a robust privacy and data protection regime designed to shield the play of everyday practice from exposure").

267. Relatedly, we might worry that this channeling would lead to less creativity: "I think creative expression is a natural byproduct of growing up in a diverse world. The more diverse the world is, the more you get exposed to different people, to different opportunities, to different places and to different challenges. And the more diverse that is, the more likely you'll be to be able to put the dots together to form something new." This was written by an AI in response to the question: "How do we become more creative?" Metz, *supra* note 176.

experiment and play in order to develop sufficient capacities to critically engage with their own projects and identities. Constraining AI advice to a greater degree will cause people to experience a greater number of spaces where they are not channeled by it, and hence, are more likely to experiment and play on their own. Without a theory about how much of this space is needed to generate this self-reflective capacity, and how much of this self-reflective capacity is needed,[268] it is difficult to say anything more with certainty.[269] But if AI advice is constrained to not operate in fields where people have conflicting views of which choices are "better," then people will have ample opportunities to reflect on deep philosophical debates and enduring moral questions. These seem like fairly fertile spaces in which to grow one's self-reflective capacity.

C. Lessons Beyond the Advice Paradigm

Even if policymakers reject or severely constrain AI advice, this Article provides important lessons. Disclosure regimes can add layers, too. They can include reason-giving, precommitments, and be savvier about the emotional context of many decisions. For example, a therapy dog may do more than any amount of statistical training to help people understand a mortality curve that is providing them with unwelcome news.[270] Similarly, many nudges could be usefully bundled together in ways that reflect the packaging of natural advice. For example, nudges that rely on social norms could be joined with efforts to increase self-efficacy and with various precommitment devices. The fact that this bundling is common in natural advice suggests that people will be receptive to it, and that it will be effective. This points to the usefulness of examining the synergies created by certain combinations

268. Needed for what? For example, this could be a question of how to maintain a well-functioning democracy, or how to flourish as an individual. *See* Cohen, *supra* note 266, at 138, 144, 146. This capacity might also be better conceptualized as a set of domain-specific capacities. For example, it is possible that you could be highly self-reflective in one domain and not in another. This might suggest the need to create multiple spaces to experiment in multiple domains.

269. Danaher, *supra* note 255, at 644 (arguing that people are nudged toward many choices and preferences by our dense social interactions with friends, family, and our broader culture, and that therefore it is "short-sighted to assume that simply because AI assistants are newer and hence more salient features of our cognitive ecology that they pose a more significant threat to our autonomy").

270. *See* Paschal Sheeran et al., *Does Heightening Risk Appraisals Change People's Intentions and Behavior? A Meta-Analysis of Experimental Studies*, 140 PSYCHOL. BULL. 511, 512 (2014) (noting the importance of emotional, not just cognitive, reactions to risk when seeking to motivate belief and behavioral change.); REGINA E. LUNDGREN & ANDREA H. MCMAKIN, RISK COMMUNICATION: A HANDBOOK FOR COMMUNICATING ENVIRONMENTAL, SAFETY, AND HEALTH RISKS 115 (5th ed. 2013) (recommending that, during a crisis, speakers start with an expression of empathy and only then proceed to discussing the basic facts).

of nudges, whether delivered simultaneously or in certain sequential orders.[271]

CONCLUSION

This Article reveals the productive potential of merging one of our most ancient technologies for the promotion of welfare—advice—with some of our most recent—AI and big data. An optimistic view of technological progress suggests that AI will soon be able to identify previously hidden patterns in our behavior to create incredibly accurate personalized advice. Analyzing our everyday practices surrounding natural advice reveals the ways that the packaging of advice can be just as important as the content. AI advice can mimic this packaging, which will significantly increase the likelihood that people will hear and heed this advice. This combination of exceedingly accurate content and emotionally cognizant packaging provides AI advice with powerful advantages over its current regulatory rivals, suggesting that it may, one day, replace mandatory disclosure and classic nudges in many policy spaces.

271. *See* Erina L. MacGeorge et al., *Advice: Expanding the Communication Paradigm*, 40 ANNALS INT'L COMM. ASS'N 213, 216 (2016) ("[T]he impact of even high quality advice messages varies as a function of interaction sequence: advice offered following emotional support and problem analysis messages is perceived as higher in quality, more facilitative of the recipient's coping, and leads to stronger implementation intention than advice that does not follow this sequential pattern.").