NORTH CAROLINA LAW REVIEW

Volume 98 | Number 4 Article 6

5-1-2020

The Extended Corporate Mind: When Corporations Use AI to Break the Law

Mihailis E. Diamantis

THE EXTENDED CORPORATE MIND: WHEN CORPORATIONS USE AI TO BREAK THE LAW*

MIHAILIS E. DIAMANTIS**

Algorithms may soon replace employees as the leading cause of corporate misconduct. For centuries, the law has defined corporate misconduct—anything from civil discrimination to criminal insider trading—in terms of employee misconduct. Today, however, breakthroughs in artificial intelligence and big data allow automated systems to make many corporate decisions, e.g., who gets a loan or what stocks to buy. These technologies introduce valuable efficiencies, but they do not remove (or even always reduce) the incidence of corporate harm. Unless the law adapts, corporations will become increasingly immune to civil and criminal liability as they transfer responsibility from employees to algorithms.

This Article is the first to tackle the full extent of the growing doctrinal gap left by algorithmic corporate misconduct. To hold corporations accountable, the law must sometimes treat them as if they "know" information stored on their servers and "intend" decisions reached by their automated systems. Cognitive science and the philosophy of mind offer a path forward. The "extended mind thesis" complicates traditional views about the physical boundaries of the mind. The thesis states that the mind encompasses any system that sufficiently assists thought, e.g., by facilitating recall or enhancing decisionmaking. For natural people, the thesis implies that minds can extend beyond the brain to include external cognitive aids, like rolodexes and calculators. This Article adapts and applies the thesis to corporate law. It proposes a doctrinal framework for extending the corporate mind to the algorithms that are increasingly integral to

** Associate Professor, University of Iowa College of Law. For invaluable feedback at various stages, I owe special thanks to Aaron Ancell, Shawn Bayern, Raff Donelson, Sean Griffith, Josh Kleinfeld, Alex Lemann, Jim Lindgren, Jeff Lipshaw, Kiel Brennan-Marquez, Micah Schwartzman, Lawrence Solum, and participants at the following workshops: the Corporate Law Workshop (University of Houston Law Center), the Business Ethics in a Digital Age Workshop (Harvard Business School), the Business Law and Technology Roundtable (Michigan State University College of Law), the Criminal Theory Workshop (Harvard Law School), the Philosophy of Law Roundtable (Southeastern Association of Law Schools), the Zicklin Center Normative Business Ethics Workshop (Wharton Business School), and the Section on Business Associations New Voices Workshop (Association of American Law Schools).

corporate thought. The law needs such an innovation if it is to hold future corporations accountable for their most serious harms.

INTRODUCTION

Marvin[1] makes investments on behalf of SciBank trying to maximize returns. Like all decent investment bankers, Marvin only purchases or sells positions after methodically collecting and weighing information about future performance.[2] One day, Marvin acquires nonpublic information that BigCo will make a bid to acquire SmallCo. Marvin's models predict that SmallCo's stock price will shoot up after BigCo announces its plan. Consequently, Marvin invests in SmallCo and makes a killing for SciBank.

Could SciBank be guilty of insider trading? If Marvin is an employee at SciBank, there are good grounds for a closer look. Crucially, since Marvin learned material, nonpublic information in the course of his employment, the law dictates that SciBank learned it as well.[3] SciBank's liability turns on the provenance of that information, i.e., whether Marvin misappropriated it[4] or received it as a tip from an insider who stood to benefit from the transaction.[5]

1. I am grateful to Aaron Ancell at the Edmond J. Safra Center for Ethics, Harvard University, for this opening conceit.

2. *See* Bernard Marr, *The Revolutionary Way of Using Artificial Intelligence in Hedge Funds*, FORBES (Feb. 15, 2019, 1:48 AM), https://www.forbes.com/sites/bernardmarr/2019/02/15/the-revolutionary-way-of-using-artificial-intelligence-in-hedge-funds-the-case-of-aidyia/#17eb640157ca [https://perma.cc/46FA-3YNJ].

3. *See* N.Y. Cent. & Hudson River R.R. Co. v. United States, 212 U.S. 481, 491 (1909); Phila., Wilmington & Balt. R.R. Co. v. Quigley, 62 U.S. (21 How.) 202, 209–10 (1858). There are some weak qualifications. *See infra* notes 173–77 and accompanying text.

4. *See* United States v. O'Hagan, 521 U.S. 642, 652 (1997).

5. *See* Dirks v. SEC, 463 U.S. 646, 659–60 (1983).

But suppose that Marvin is not a human being at all; suppose "Marvin" is one of many algorithmic trading programs in use today.[6] Though Marvin may still have acquired the information illegitimately (through misappropriation or an improper tip), the insider trading inquiry immediately aborts. Since Marvin is not a human employee, the law has no way to say SciBank knew the information about BigCo's planned acquisition. That makes liability for insider trading a nonstarter.[7]

The SciBank hypothetical is neither futuristic nor idiosyncratic.[8] Advanced algorithms utilizing big data and artificial intelligence are rapidly reshaping every corner of modern business.[9] Experts predict that corporate reliance on digital automation will increase exponentially over the coming years.[10] Algorithms are taking over human functions throughout the corporate hierarchy, from the lowest-level operations—like the systems running

6. *See* Marr, *supra* note 2.

7. *See* 17 C.F.R. § 240.10b5-1(a) (2019).

8. Computer scientists proved years ago that algorithms could teach themselves to manipulate markets. *See* Tom C.W. Lin, *The New Market Manipulation*, 66 EMORY L.J. 1253, 1284–86, 1292–94 (2017) (discussing how AI can learn to engage in pump-and-dump manipulation); Enrique Martínez-Miranda, Peter McBurney & Matthew J. Howard, *Learning Unfair Trading: A Market Manipulation Analysis from the Reinforcement Learning Perspective*, 2016 IEEE CONF. ON EVOLVING & ADAPTIVE INTELLIGENT SYSS. 103, 108; Ben Van Lier, *From High Frequency Trading to Self-Organizing Moral Machines*, 7 INT'L J. TECHNOETHICS 34, 34 (2016) (noting that AI has become increasingly autonomous in regards to its role in the financial sector); Michael P. Wellman & Uday Rajan, *Ethical Issues for Autonomous Trading Agents*, 27 MINDS & MACHINES 609, 614 (2017); Renato Zamagna, *The Future of Trading Belong to Artificial Intelligence*, MEDIUM (Nov. 15, 2018), https://medium.com/datadriveninvestor/the-future-of-trading-belong-to-artificial-intelligence-a4d5887cb677 [http://perma.cc/X5HW-VK2H]; *see also* Council Regulation 596/2014 of 16 Apr. 2014, On Market Abuse (Market Abuse Regulation) and Repealing Directive 2003/6/EC of the European Parliament and of the Council and Commission Directives 2003/124/EC, 2003/125/EC and 2004/72/EC, 2014 O.J (L. 173) 1; Council Directive 2014/57/EU of 16 Apr. 2014, On Criminal Sanctions for Market Abuse (Market Abuse Directive), 2014 O.J. (L. 173) 179.

9. *See* Edward L. Pittman, *Quantitative Investment Models, Errors, and the Federal Securities Laws*, 13 N.Y.U. J.L. & BUS. 633, 643–44 (2017) (discussing near universal use of algorithms and quantitative tools in investment management); H. James Wilson & Paul R. Daugherty, *Collaborative Intelligence: Humans and AI Are Joining Forces*, HARV. BUS. REV., July–Aug. 2018, at 114, 116–18 (noting the rise of AI and emphasizing the necessity of collaboration); Dan Wellers, Timo Elliott & Markus Noga, *8 Ways Machine Learning Is Improving Companies' Work Processes*, HARV. BUS. REV. (May 31, 2017), https://hbr.org/2017/05/8-ways-machine-learning-is-improving-companies-work-processes [http://perma.cc/SP5Q-FZ9W].

10. *See* SAM RANSBOTHAM ET AL., RESHAPING BUSINESS WITH ARTIFICIAL INTELLIGENCE 14 (2017), https://www.bcg.com/Images/Reshaping%20Business%20with%20Artificial%20Intelligence_tcm9-177882.pdf [https://perma.cc/9FDZ-L4SX]; Wellers et al., *supra* note 9.

Amazon's box-packing bots[11]—to the highest—like "Vital," the algorithm appointed to the board of Deep Knowledge Ventures.[12]

Algorithms promise to make corporations more efficient[13] and (perhaps)[14] more objective,[15] but they do not remove (or even always reduce)[16] the possibility that things will sometimes go awry.[17] Indeed, the speed and geographic reach of algorithmic processes means that when things *do* go wrong, they can go *really* wrong for a lot of people in a lot of places at once.[18] Real-life

11. Amazon uses robots to receive and automatically package many customer orders. Jeffrey Dastin, *Exclusive: Amazon Rolls out Machines That Pack Orders and Replace Jobs*, REUTERS (May 13, 2019, 6:08 AM), https://www.reuters.com/article/us-amazon-com-automation-exclusive/exclusive-amazon-rolls-out-machines-that-pack-orders-and-replace-jobs-idUSKCN1SJ0X1 [https://perma.cc/JY99-62T5].

12. Vital is credited with helping Deep Knowledge Ventures avoid bankruptcy by more logically evaluating potential investments. Nicky Burridge, *Artificial Intelligence Gets a Seat in the Boardroom*, NIKKEI ASIA REV. (May 10, 2017), https://asia.nikkei.com/Business/Artificial-intelligence-gets-a-seat-in-the-boardroom [https://perma.cc/6SDR-5XGN]. For a general discussion on AI in board rooms, see Mark Fenwick, Wulf A. Kaal & Erik P.M. Vermeulen, *The "Unmediated" and "Tech-Driven" Corporate Governance of Today's Winning Companies*, 16 N.Y.U. J.L. & BUS. 75, 114–15 (2019) ("Additionally, artificial intelligence may soon become an integral part of board decision-making. It is conceivable that future boards will include a seat for an artificial intelligent board member with voting authority. In the not too distant future, it seems feasible that artificial intelligence will have an independent board seat and may be trusted to make smarter data-driven choices than humans."); Sergio Gramitto, *The Technology and Archaeology of Corporate Law* 33–40 (Cornell Legal Studies, Research Paper No. 18-40, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3232816 [https://perma.cc/K4HQ-E6EP] ("While artificial intelligence can be used in boardrooms simply to assist human directors with making decisions, it also has the potential to replace human directors entirely.").

13. *See* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 65 (2019) ("Algorithms hold tremendous value. Big data promises significant benefits to the economy, allowing customers to find and sort products more quickly, which in turn lowers search costs.").

14. *See* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 142–45 (2016) (discussing the rise of algorithms as well as the benefits and issues created by things such as FICO and e-scores); Paul Schwartz, *Data Processing and Governance Administration: The Failure of the American Response to the Computer*, 43 HASTINGS L.J. 1321, 1342 (1992); Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. (Apr. 1, 2013), https://hbr.org/2013/04/the-hidden-biases-in-big-data [https://perma.cc/5C9K-XYM9].

15. *See* Matthew Adam Bruckner, *The Promise and Perils of Algorithmic Lenders' Use of Big Data*, 93 CHI.-KENT L. REV. 3, 23 (2018); Jason Kreag, *Prosecutorial Analytics*, 94 WASH. U. L. REV. 771, 785 (2017); Chris Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, WIRED (June 23, 2008, 12:00 PM), https://www.wired.com/2008/06/pb-theory/ [https://perma.cc/Q5QJ-EH42].

16. Cade Metz, *Is Ethical A.I. Even Possible?*, N.Y. TIMES (Mar. 1, 2019), https://www.nytimes.com/2019/03/01/business/ethics-artificial-intelligence.html [https://perma.cc/YMG8-JPMX (dark archive)].

17. *See* Bruckner, *supra* note 15, at 6; Mark A. Geistfeld, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, 105 CALIF. L. REV. 1611, 1614–16, 1620 (2017) ("Autonomous vehicles will not be perfectly safe."); Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. CHI. L. REV 1311, 1313 (2019) ("As robotics and artificial intelligence (AI) systems increasingly integrate into our society, they will do bad things.").

18. *See* EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES, at iii (2014), https://obamawhitehouse.archives.gov/sites/default

examples of corporate algorithmic harm that merit a searching liability inquiry include[19]:

- A lender's automated platform approving mortgages in a fashion that has a discriminatory racial impact but might also have a business justification.[20]
- Competing retailers' pricing algorithms setting prices at matching, super-competitive levels.[21]
- A delivery company's self-driving truck striking a jaywalking pedestrian.[22]

Not long ago, corporations relied on human employees to carry out each of these functions. Today, many corporations use algorithms to approve loans, set prices, and transport goods.[23]

The move toward automation does not alter the fact that discrimination, price fixing, and reckless driving leave victims in their wake.[24] These victims,

/files/docs/big_data_privacy_report_may_1_2014.pdf [https://perma.cc/QWG5-USYP] (describing the potential of algorithms to undermine longstanding civil rights protections).

19. A growing scholarly literature discusses others. *See, e.g.*, Katyal, *supra* note 13, at 56.

20. *See generally* Robin Nunn, *Discrimination and Algorithms in Financial Services: Unintended Consequences of AI*, LEXOLOGY (Mar. 6, 2018), https://www.lexology.com/library/detail.aspx?g=ecdd186d-29eb-4821-8161-89b93cf1ef31 [https://perma.cc/9SHT-3Q6V (dark archive)] (discussing the importance of racially sensitive implementation of AI). For a similar example involving hiring ads, see Esha Bhandari & Rachel Goodman, *ACLU Challenges Computer Crimes Law That Is Thwarting Research on Discrimination Online*, ACLU (June 29, 2016, 10:00 AM), https://www.aclu.org/blog/racial-justice/race-and-economic-justice/aclu-challenges-computer-crimes-law-thwarting-research [https://perma.cc/838L-YZEE]. *See also* Mikella Hurley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 YALE J.L. & TECH. 148, 194 (2016) (discussing the challenges of such cases).

21. *See* Emilio Calvano et al., *Artificial Intelligence, Algorithmic Pricing, and Collusion*, VOX (Feb. 3, 2019), https://voxeu.org/article/artificial-intelligence-algorithmic-pricing-and-collusion [https://perma.cc/6WPN-SHXF]; Greg Rosalsky, *When Computers Collude*, NPR (Apr. 2, 2019, 7:30 AM), https://www.npr.org/sections/money/2019/04/02/708876202/when-computers-collude [https://perma.cc/ZE2T-6UUE]; *see also* Maurice E. Stucke & Ariel Ezrachi, *Two Artificial Neural Networks Meet in an Online Hub and Change the Future (of Competition, Market Dynamics and Society)* 52 (U. Tenn. Legal Stud., Research Paper No. 323, July 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2949434 [https://perma.cc/6RNH-YSRN] (discussing "algorithmic price optimization").

22. Daisuke Wakabayashi, *Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam*, N.Y. TIMES (Mar. 19, 2018), https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html [https://perma.cc/MWD5-D8FK (dark archive)].

23. *See generally* Ellen Ruppel Shell, *AI and Automation Will Replace Most Human Workers Because They Don't Have To Be Perfect—Just Better Than You*, NEWSWEEK (Nov. 20, 2018), https://www.newsweek.com/2018/11/30/ai-and-automation-will-replace-most-human-workers-because-they-dont-have-be-1225552.html [https://perma.cc/BG83-TAT7] (describing how mega-tech companies are overwhelmingly automating their workforce).

24. For a detailed discussion of the harms of algorithmic discrimination, see O'NEIL, *supra* note 14, at 13 ("[G]oing to college, borrowing money, getting sentenced to prison, or finding and holding a job. All of these life domains are increasingly controlled by secret [big data algorithms] wielding arbitrary punishments.").

or the state on their behalf, should have as clear a path to justice as their counterparts a decade ago. In cases of algorithmic misconduct, it is particularly important that the path hold open the possibility of corporate liability. As corporations replace employees with algorithms, corporate liability becomes the *only* means of redress. Employees are accountable for their own misconduct, whether on or off the job.[25] Algorithms, however, are not subject to suit.[26]

The problem is that the law is not equipped to address corporate liability when the "thinking" behind corporate misconduct has been offloaded to automated systems.[27] Under current law, corporate liability in each of the above cases would require (and, as I assume below, should require)[28] evidence of a culpable corporate mental state: purpose (to discriminate),[29] knowledge (of competitors' prices),[30] or recklessness (in operating a vehicle).[31] The legal doctrine for attributing mental states to corporations—respondeat superior—defines corporate mental states in terms of employee mental states.[32] Variants of respondeat superior—like the collective knowledge doctrine, which allows courts to aggregate employee knowledge,[33] and the control group test, which restricts respondeat superior to higher ranking corporate employees[34]—only

25. *See generally* V.S. Khanna, *Corporate Criminal Liability: What Purpose Does It Serve?*, 109 HARV. L. REV. 1477, 1489–90 (1996) (discussing the doctrine of respondeat superior under which a corporation can also be held liable for an individual employee's actions).

26. United States v. Athlone Indus., Inc., 746 F.2d 977, 979 (3d Cir. 1984) (stating that "[r]obots cannot be sued"); Ugo Pagallo, *Killers, Fridges, and Slaves: A Legal Journey in Robotics*, 26 AI & SOC'Y 347, 349 (2011) ("The common legal standpoint excludes robots from any kind of criminal responsibility . . . .").

27. This parallels another philosophical problem: How can data stored digitally have "meaning"? *See generally* Lawrence B. Solum, *Artificial Meaning*, 89 WASH. L. REV. 69, 69–70 (2014) (discussing "the concept of artificial meaning, meanings produced by entities other than individual natural persons").

28. *See infra* Part I.

29. Discrimination cases can be brought even in the absence of purpose if they allege a disparate impact. *See, e.g.*, *CFPB and DOJ Order Ally To Pay $80 Million to Consumers Harmed by Discriminatory Auto Loan Pricing*, CFPB NEWSROOM (Dec. 20, 2013), https://www.consumerfinance.gov/about-us/newsroom/cfpb-and-doj-order-ally-to-pay-80-million-to-consumers-harmed-by-discriminatory-auto-loan-pricing/ [https://perma.cc/KC6K-3BKE]. Where there is a possible business justification for the disparate treatment (as there generally will be in cases of algorithmic discrimination), the case becomes much more difficult in the absence of proof of purpose. Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 711–12, 726 (2016); Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 535 (2018).

30. *See* Sherman Antitrust Act, ch. 647, 26 Stat. 209 (1890) (codified as amended at 15 U.S.C. §§ 1–7 (2018)); United States v. Wise, 370 U.S. 405, 416 (1962) ("[A] corporate officer is subject to prosecution under § 1 of the Sherman Act whenever he knowingly participates in effecting the illegal contract, combination, or conspiracy . . . .").

31. *See, e.g.*, IOWA CODE ANN. § 321.277 (Westlaw through 2019 legislation) (defining the offense of "reckless driving").

32. RESTATEMENT (THIRD) OF AGENCY § 2.04 cmt. b (AM. LAW INST. 2006). I discuss these weak limits below. *See infra* Part IV.

33. *See* United States v. Bank of New England, N.A., 821 F.2d 844, 856 (1st Cir. 1987).

34. *See* MODEL PENAL CODE § 2.07(1)(c) (AM. LAW INST. 2018).

reinforce the current legal fact that *corporate* mental states must derive from *employee* mental states. When corporations misbehave through their employees, respondeat superior produces relatively straightforward liability determinations.[35] But when corporations misbehave through their algorithms in ways that, from the outside, look just as purposeful, knowing, or reckless as the misbehavior carried out by human employees, current liability doctrines do not apply.[36]

In a commercial world increasingly run on silicon, it is surprising that the law's understanding of the corporate mind is still tied to a prehistoric lump of grey organic matter. A corporation like JPMorgan has at its fingertips server data that literally exceeds the storage capacity of, on some calculations, 390,000 human brains.[37] Its processors analyze that information, on some estimates, 10,000,000 times faster than any human could.[38] If the information and any conclusions drawn from it do not pass through a human employee's brain, they form no part of the law's present conception of the corporate mind.

The current state of the law is troubling because it all but guarantees that corporations will become increasingly immune to liability as their operations require less and less human intervention.[39] For example, lacking a theory of liability, prosecutors declined to file charges against Uber when one of its self-driving cars struck and killed a pedestrian in Arizona.[40] The legal loophole left by respondeat superior incentivizes an unpalatable form of corporate

35. The results of these straightforward determinations are not particularly compelling. *See* Mihailis E. Diamantis, *Corporate Criminal Minds*, 91 NOTRE DAME L. REV. 2049, 2056–58 (2016) [hereinafter Diamantis, *Corporate Criminal*] (critiquing respondeat superior).

36. Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217, 1234 (2017).

37. *See* Dakin Campbell, *Meet the JPMorgan Banker with No Technical Expertise Who's Now in Charge of the Biggest Data Projects on Wall Street*, BUS. INSIDER (June 12, 2019), https://www.businessinsider.com/rob-casper-jpmorgan-data-head-profile-managing-financial-account-info-2019-5 [https://perma.cc/6BDR-W9SA]; Forrest Wickman, *Your Brain's Technical Specs*, SLATE (Apr. 24, 2012, 8:18 PM), https://slate.com/technology/2012/04/north-koreas-2-mb-of-knowledge-taunt-how-many-megabytes-does-the-human-brain-hold.html [https://perma.cc/6SFR-AV4D].

38. Liqun Luo, *Why Is the Human Brain So Efficient?*, NAUTILUS (Apr. 12, 2018), http://nautil.us/issue/59/connections/why-is-the-human-brain-so-efficient [https://perma.cc/S2LD-44UA].

39. A related set of issues arise as governments use algorithms to perform diverse functions, such as evaluating social security benefits claims and making parole determinations. *See generally* Ronald Bailey, *Welcoming Our New Algorithmic Overlords?*, REASON (July 8, 2016, 1:30 PM), https://reason.com/2016/07/08/welcoming-our-new-algorithmic-overlords-2 [https://perma.cc/4DP8-SMU6] (discussing the use of algorithms in governmental and judicial decisionmaking).

40. Angie Schmitt, *Uber Got off the Hook for Killing a Pedestrian with Its Self-Driving Car*, STREETSBLOG USA (Mar. 8, 2019), https://usa.streetsblog.org/2019/03/08/uber-got-off-the-hook-for-killing-a-pedestrian-with-its-self-driving-car/ [https://perma.cc/3V4D-K2UV] (speculating that "tech companies won't be punished for taking egregious risks with their untested technology even when the worst happens.").

gamesmanship. Corporations keen to manage their liabilities[41] will seek the safe haven of algorithmic misconduct rather than chance liability for misconduct by human employees. By not providing a solution, the law incentivizes corporations to accelerate their embrace of automation. This dynamic exacerbates the risk that corporations will turn to algorithms prematurely, before the technology has been sufficiently tested for socially responsible use.[42]

The law needs a framework for extending its conception of the corporate mind beyond the employees whose shoes algorithms are coming to fill. Only then could the law develop reliable doctrines for evaluating whether corporations that misbehave through algorithms nonetheless satisfy the mental state elements of liability. Psychologists and philosophers have recently addressed a related set of issues about the human mind. They argue that the traditional understanding of the human mind as limited by the boundaries of the skull is too restrictive.[43] The so-called "extended mind thesis" states that the human mind reaches beyond the brain to encompass external cognitive aids—such as diaries or cellphones—that help the brain do its work.[44] If a person can as easily "recollect" a phone number by checking her phone's memory bank as by checking her neurological memory bank, her mind may, according to the thesis, extend to aspects of her phone.[45] This Article adapts, with appropriate modifications, the extended mind thesis to the corporate context. It argues that the law could and should recognize that corporate minds extend to algorithms fulfilling roles that were once occupied only by human employees. By extending the corporate mind in this way, the law could bring corporate accountability into the twenty-first century.

The proposal developed below is targeted at judges and/or prosecutors. The liability framework applicable to corporations was largely derived from

41. Corporations are assumed to want to do this. *See* Cindy R. Alexander & Mark A. Cohen, *The Causes of Corporate Crime: An Economic Perspective*, *in* PROSECUTORS IN THE BOARDROOM: USING CRIMINAL LAW TO REGULATE CORPORATE CONDUCT 11, 14–15, 17 (Anthony S. Barkow & Rachel E. Barkow eds., 2011).

42. Microsoft President and Chief Legal Officer Brad Smith has remarked that "[w]e don't want to see a commercial race to the bottom" and stated that "[l]aw is needed." Metz, *supra* note 16; *see also* Frank Pasquale, *Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society*, 78 OHIO ST. L.J. 1243, 1244–45 (2017) (describing North Carolina's attempt to prohibit legal software manufacturers from attaining a lower standard of liability than attorneys and the Federal Trade Commission and Department of Justice's criticism of North Carolina's decision).

43. *See* Andy Clark & David Chalmers, *The Extended Mind*, 58 ANALYSIS 7, 8–10 (1998) (introducing the extended mind thesis, which the authors refer to as "active externalism").

44. *See* Marc Jonathan Blitz, *Freedom of Thought for the Extended Mind: Cognitive Enhancement and the Constitution*, 2010 WIS. L. REV. 1049, 1055–56 (describing Clark and Chalmers's theory as the "extended mind"); Clark & Chalmers, *supra* note 43, at 12–14 (discussing an example involving a notebook).

45. *See* Clark & Chalmers, *supra* note 43, at 12–14. As I discuss further below, the extended mind thesis does have some limits. *See infra* Section IV.B.

common law,[46] was introduced to corporate law, and then expanded through judicial activity.[47] As such, judicial decisions are the likeliest point of evolution for existing doctrine.[48] In criminal law, prosecutors also have an important role to play as gatekeepers for corporate liability. Lacking a theory of liability, prosecutors may, as with Uber's reckless driving, decline to bring charges. Alternatively, prosecutors may address the obvious need for criminal liability in cases of algorithmic corporate misconduct on their own by coercing out-of-court agreements.[49] In the absence of a principled approach, prosecutors are likely to exacerbate rule-of-law[50] and accountability[51] concerns that others have raised about how prosecutors resolve suspected cases of corporate crime. The solution offered below could provide some guidance.

This Article begins by clarifying important aspects of its methodology (Part I) and offering a focusing hypothetical scenario (Part II). In its chief substantive contribution, this Article introduces the extended mind thesis (Part III) and shows in detail how to adapt this concept as a doctrine for addressing algorithmic corporate misconduct (Part IV). Finally, this Article concludes by reflecting on broader implications of the extended corporate mind.

I. A MINIMALLY INVASIVE METHOD

There are two typical tools for solving legal problems: sledgehammers and scalpels. Sledgehammers are for addressing basic structural defects in the law when the only path forward is wholesale reform. Their basic function is to demolish and rebuild. This Article's more modest ambition is to use a scalpel.

46. *See* Packard Motor Car Co. v. NLRB, 330 U.S. 485, 489 (1947) (referring to "the ancient maxim of the common law, *respondeat superior*," which, "[e]ven without special statutory provision[,] . . . would apply to many relations").

47. *See* Samuel W. Buell, *The Blaming Function of Entity Criminal Liability*, 81 IND. L.J. 473, 474–75 (2006) ("The law in this area had a weak start nearly a century ago when common law courts, looking to expand available means for regulating business enterprises, imported respondeat superior liability from tort law into the criminal law, but without serious theoretical analysis.").

48. The only serious efforts at reform have come from judges. *See, e.g.*, United States v. Bank of New England, N.A., 821 F.2d 844, 855 (1st Cir. 1987) (upholding the trial court's jury instructions to impute the "collective knowledge" of all employees to the employer on an issue of respondeat superior liability).

49. *See* Brandon L. Garrett, *Collaborative Organizational Prosecution*, *in* PROSECUTORS IN THE BOARDROOM: USING CRIMINAL LAW TO REGULATE CORPORATE CONDUCT, *supra* note 41, at 154, 157 ("[P]rosecutors typically defer prosecution or agree not to prosecute if a firm will enter into an agreement.").

50. *See generally* Jennifer Arlen, *Prosecuting Beyond the Rule of Law: Corporate Mandates Imposed Through Deferred Prosecution Agreements*, 8 J. LEGAL ANALYSIS 191 (2016) (arguing that the wide variance between the terms prosecutors impose through deferred prosecution agreements ("DPAs") and non-prosecution agreements ("NPAs") violates the rule of law).

51. *See* Lisa Kern Griffin, *Inside-Out Enforcement*, *in* PROSECUTORS IN THE BOARDROOM: USING CRIMINAL LAW TO REGULATE CORPORATE CONDUCT, *supra* note 41, at 110, 110 ("DPAs are less visible than adjudication, which detracts from both the coherence of the government's enforcement strategy and the accountability of prosecutors.").

It aims to solve the problem of algorithmic corporate misconduct by making the smallest possible doctrinal incision. What surgical intervention may sacrifice in terms of grandiose vision, it makes up in terms of feasibility. It does this by leveraging existing frameworks and doctrines—in this case, of corporate liability—to address the problem of algorithmic misconduct. Small changes are more likely to get real-life traction because they tend to be more palatable to lawmakers than large changes.[52] This feature is the present proposal's distinctive advantage.

This Article is about corporate criminal and civil liability for algorithmic misconduct. True to its surgical aspirations, the remainder of this part flags some related but ultimately tangential issues. I mean to leave the law on these issues undisturbed. In doing so, I hope to solve the problem of algorithmic corporate misconduct while dodging several broader problems that have interested other theorists.

Some scholars believe that corporate culpability is such a nonsensical notion[53] that the law would be better without it.[54] They may be right. After all, corporations are just fictional agents.[55] Perhaps the best solution to algorithmic corporate misconduct would begin by critically engaging the law's commitment to corporate culpability.

In accord with its surgical ambitions, this Article does not question whether corporations can be (or should be) legally accountable. Instead, it aims, so far as possible, to leave this corner of the law undisturbed. As I have argued elsewhere, the law's general conception of corporations as responsible actors is psychologically sustainable[56] and often makes good policy sense.[57] More to the

52. *See* GUIDO CALABRESI, A COMMON LAW FOR THE AGE OF STATUTES 3–4 (1982); Boris I. Bittker, *Interpreting the Constitution: Is the Intent of the Framers Controlling? If Not, What Is?*, 19 HARV. J.L. & PUB. POL'Y 9, 51–52 (1995); Cynthia R. Farina, *Faith, Hope, and Rationality or Public Choice and the Perils of Occam's Razor*, 28 FLA. ST. U. L. REV. 109, 110–11 (2000) (noting that "public choice is 'appealing in its parsimoniousness'" and that "[e]ven if we had the ability to dismantle the entire national regulatory apparatus, we have neither the will nor the desire to do so" (quoting Steven P. Corley, *Public Interested Regulation*, 28 FLA. ST. U.L. REV. 7, 15 (2000))); Saul Levmore, *Interest Groups and the Problem with Incrementalism*, 158 U. PA. L. REV. 815, 816–17 (2010) ("Leading commentators encourage incrementalism. Most of the encouragement is directed at judges, but the arguments used in favor of incrementalism are equally applicable to regulators and legislators." (footnote omitted)).

53. *See, e.g.*, Amy J. Sepinwall, *Guilty by Proxy: Expanding the Boundaries of Responsibility in the Face of Corporate Crime*, 63 HASTINGS L.J. 411, 428 (2012) (arguing that corporations cannot possess moral agency because they have no capacity for moral emotions).

54. *See, e.g.*, John Hasnas, *The Centenary of a Mistake: One Hundred Years of Corporate Criminal Liability*, 46 AM. CRIM. L. REV. 1329, 1329 (2009).

55. *See* Sierra Club v. Morton, 405 U.S. 727, 742–43 (1972) (Douglas, J., dissenting) ("The ordinary corporation is a 'person' for purposes of the adjudicatory processes . . . .").

56. *See* Gerhard O.W. Mueller, *Mens Rea and the Corporation: A Study of the Model Penal Code Position on Corporate Criminal Liability*, 19 U. PITT. L. REV. 21, 40–41 (1957).

57. *See* Diamantis, *Corporate Criminal*, *supra* note 35, at 2052–53 (explaining why applying a mens rea to corporations makes sense).

point, scrapping corporate culpability is a practical nonstarter. The public broadly supports holding corporations civilly and criminally accountable.[58] Consequently, the law of corporate culpability is socially entrenched and politically bulletproof. Furthermore, the legal edifice built around corporate culpability is centuries old[59] and growing.[60] Taking it apart would require a massive jurisprudential undertaking.

No feature of the law's framework for corporate culpability is more integral than the fiction of corporate personhood.[61] According to this sociolegal construct, corporations fit into the law's liability mechanisms just as other "people" do.[62] Any statute that defines a civil or criminal violation simultaneously sets out elements of liability for both individuals and corporations.[63] Whatever action a statute requires, it requires of real and fictional people alike.[64] Where, as is most often the case,[65] liability also turns on the person having a concurrent mental state, both individual and corporate

58. *See* Miriam H. Baer, *Choosing Punishment*, 92 B.U. L. REV. 577, 612 (2012).

59. The doctrine was introduced to U.S. civil law in the mid-nineteenth century. *See* Phila. & Reading R.R. Co. v. Derby, 55 U.S. (14 How.) 468, 486–87 (1852) (establishing a railroad's liability for the negligence of its employee through the doctrine of respondeat superior). Fifty years later, it came to U.S. criminal law. *See* N.Y. Cent. & Hudson River R.R. Co. v. United States, 212 U.S. 481, 494–96 (1909) (holding that finding a corporation guilty of committing a crime does not violate the Constitution).

60. *See* Sara Sun Beale, *A Response to the Critics of Corporate Criminal Liability*, 46 AM. CRIM. L. REV. 1481, 1482 (2009).

61. *See* David M. Uhlmann, *The Pendulum Swings: Reconsidering Corporate Criminal Prosecution*, 49 U.C. DAVIS L. REV. 1235, 1246 (2016) (acknowledging that corporate prosecution is based on the legal fiction of corporations' personhood under the law); *see also Corporation*, BLACK'S LAW DICTIONARY (11th ed. 2019) (defining a corporation as an entity "having authority under law to act as a single person").

62. *See, e.g.*, Dictionary Act of 1871, ch. 71, § 2, 16 Stat. 431, 431 (current version at 1 U.S.C. § 1 (2018)) (defining "person" to include "corporation"); Uhlmann, *supra* note 61, at 1246. There is an important distinction between "personhood"—which applies to entities that have rights and responsibilities—and "peoplehood"—which applies to persons that have minds. While people are persons, the reverse is not necessarily true, except where the responsibility at issue in personhood is legal liability that requires mental states. Given this Article's focus on legal liability of precisely that sort, I use "people" and "persons" interchangeably to refer to entities that have rights, responsibilities, and minds.

63. *See* 1 U.S.C. § 1 (2018).

64. For example, when a statute defines the crime that applies to "[w]hoever . . . corruptly persuades another person . . . to . . . destroy, mutilate, or conceal an object with intent to impair the object's integrity or availability for use in an official proceeding," 18 U.S.C. § 1512(b)(2) (2018), there is no question that "whoever" includes corporations, *see* United States v. Arthur Andersen, LLP, 374 F.3d 281, 284 (5th Cir. 2004) (upholding accounting firm's conviction for obstruction of justice), *rev'd on other grounds*, 544 U.S. 696 (2005).

65. *See* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM ch. 4, scope note (AM. LAW INST. 2010) (noting that strict liability is generally limited to torts involving abnormally dangerous activity, possession of animals, and products liability); RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 1 (AM. LAW. INST. 1998) (explaining the development of strict liability in products liability cases); 22 C.J.S. *Criminal Law* § 39 (2006) ("Strict liability statutes remain the exception in our criminal law system, not the rule, and have a generally disfavored status.").

defendants must satisfy that mental state to be liable.[66] Some theorists think that because corporations are not really people, the law should abandon its pretense to the contrary.[67] These theorists argue that the law's present reliance on corporate mental states makes for ineffective justice[68] and suboptimal prevention.[69] Removing any reference to corporate mental states would convert all corporate liability into strict liability.

This Article is not about the advisability of the law's fiction of corporate personhood. The law can indulge this fiction regardless of whether[70] or not[71] corporations really are people. Indeed, the fact that the law self-consciously invokes a "fiction" of corporate personhood is itself a concession that corporations are not really people and do not really have mental states.[72] As I have argued elsewhere, this pretense helps the law accomplish its basic goals. The fiction that corporations can have culpable mental states helps the law identify truly reprehensible corporate conduct (as opposed to merely harmful conduct) for distinctive treatment.[73] Additionally, despite all the sophisticated economic theory endorsing strict corporate liability,[74] that approach often

66. Khanna, *supra* note 25, at 1489 ("[Among the] requirements [that] must be met in order to impose liability on a corporation . . . a corporate agent must have committed an illegal act (the actus reus) with the requisite state of mind (the mens rea)."); W. Robert Thomas, *Incapacitating Criminal Corporations*, 72 VAND. L. REV. 905, 914–15 (2019) ("[T]he practice of holding commercial corporations criminally responsible for general intent crimes as well as specific intent crimes—crimes for which there exists a proscribed action (actus reus) pursued concurrently with a prescribed attitude (mens rea)—took hold around the turn of the twentieth century.").

67. *See* John S. Baker, Jr., *Reforming Corporations Through Threats of Federal Prosecution*, 89 CORNELL L. REV. 310, 349–53 (2004) (describing how corporate defendants' differences from human defendants place them at a disadvantage).

68. *See* Albert W. Alschuler, *Two Ways To Think About the Punishment of Corporations*, 46 AM. CRIM. L. REV. 1359, 1392 (2009) [hereinafter Alschuler, *Two Ways To Think*] (arguing that two possible legal analogies for corporate punishment—deodand, "the punishment of animals and inanimate objects that produced harm," or frankpledge, "the punishment of all members of a group when one member of the group has avoided apprehension for a crime"—show its absurdity).

69. *See* Daniel R. Fischel & Alan O. Sykes, *Corporate Crime*, 25 J. LEGAL STUD. 319, 320–21 (1996) (arguing that imposing criminal liability on corporations results in overdeterrence).

70. *See* CHRISTIAN LIST & PHILIP PETTIT, GROUP AGENCY: THE POSSIBILITY, DESIGN, AND STATUS OF CORPORATE AGENTS 1–2 (2011) (arguing that corporations are agents).

71. MAX WEBER, ECONOMY AND SOCIETY: A NEW TRANSLATION 89–90 (Keith Tribe ed. & trans., 2019) (arguing that corporations are not agents).

72. *See, e.g.*, Int'l Shoe Co. v. Washington, 326 U.S. 310, 316 (1945) ("[T]he corporate personality is a fiction, although a fiction intended to be acted upon as though it were a fact . . . .").

73. Diamantis, *Corporate Criminal*, *supra* note 35, at 2063–64.

74. *See, e.g.*, Fischel & Sykes, *supra* note 69, at 328 ("[W]here the agent's crime is properly viewed as a cost of corporate activity, it seems appropriate that the corporation should bear 'strict' liability for the social cost of the crime . . . . The conventional justification for strict vicarious liability . . . [is] the importance of cost internalization."); Richard A. Posner, *An Economic Theory of the Criminal Law*, 85 COLUM. L. REV. 1193, 1222, 1228–29 (1985) ("In effect we introduce a degree of strict liability into criminal law as into tort law when a change in activity level is an efficient method of avoiding a social cost.").

results in inefficient overdeterrence[75] and overinvestment in compliance.[76] Regardless of its (de)merits, abandoning the fiction of corporate personhood would require a sweeping reimagination of current corporate law. It would therefore be contrary to the surgical approach adopted here. This Article situates itself within the law's fictions of corporate personhood and corporate mentality. Its goal is to find a sensible extension of that fiction to accommodate cases of algorithmic corporate misconduct.[77]

Once again, this Article will not claim, implicitly or otherwise, that corporations *actually do* have mental states. Instead, it explores what it would mean as a conceptual matter for corporations to think things. Since the law is committed to a fiction in which corporations have mental states, what is it also committed to? What shape can or must that fiction take? As shown below, the answers to those questions could hold the solution to the problem of algorithmic corporate misconduct.

Lastly, this Article will not follow the lead of scholars in law,[78] computer science,[79] and business ethics[80] who propose addressing algorithmic misconduct head-on by holding the algorithms themselves liable. That approach is deeply controversial. It is far from clear that algorithms presently do, or ever could,[81] satisfy the conditions of personhood and culpability.[82] Even theorists who

75. *See* United States v. U.S. Gypsum Co., 438 U.S. 422, 440–43 (1978); *Developments in the Law—Corporate Crime: Regulating Corporate Behavior Through Criminal Sanctions*, 92 HARV. L. REV. 1227, 1270 (1979).

76. Mihailis E. Diamantis, *Functional Corporate Knowledge*, 61 WM. & MARY L. REV. 319, 348–49 (2019) [hereinafter Diamantis, *Functional Corporate*] (explaining how the collective knowledge approach leads to extreme focus on compliance).

77. By situating itself within the legal fiction of corporate personhood, this Article need not make any controversial assumptions about the true metaphysics of corporations.

78. *See* GABRIEL HALLEVY, LIABILITY FOR CRIMES INVOLVING ARTIFICIAL INTELLIGENCE SYSTEMS 27–28 (2015); Steven J. Frank, *Tort Adjudication and the Emergence of Artificial Intelligence Software*, 21 SUFFOLK U. L. REV. 623, 624–25 (1987); Christina Mullitan, *Revenge Against Robots*, 69 S.C. L. REV. 579, 579–80 (2018).

79. *See* Fahad Alaieri & André Vellino, *Ethical Decision Making in Robots: Autonomy, Trust and Responsibility*, *in* SOCIAL ROBOTICS: 8TH INTERNATIONAL CONFERENCE 159, 159 (Arvin Agah et al. eds., 2016) ("[N]on-predictability and autonomy may confer a greater degree of responsibility to the machine . . . ."); Luciano Floridi & J.W. Sanders, *On the Morality of Artificial Agents*, 14 MINDS & MACHINES 349, 373–74 (2004); Gabriel Hallevy, *Unmanned Vehicles: Subordination to Criminal Law Under the Modern Concept of Criminal Liability*, 21 J. L. INFO. & SCI. 200, 200 (2011).

80. *See* Nicholas Diakopoulos & Sorelle Friedler, *How To Hold Algorithms Accountable*, MIT TECH. REV. (Nov. 17, 2016), https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/ [https://perma.cc/4B7Z-X6MF (dark archive)].

81. *See* JOHN SEARLE, MINDS, BRAINS AND SCIENCE 30–31 (1984) (arguing that computers cannot think); Thomas C. King et al., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, 26 SCI. & ENGINEERING ETHICS 89, 95, 102 (2019) (noting both that "the idea that an [algorithm] can act voluntarily is contentious" and that an artificial agent "cannot itself meet the *mens rea* requirement [of a crime]").

82. *See generally* JOHN CHIPMAN GRAY, THE NATURE AND SOURCES OF THE LAW 27–52 (1909) (discussing legal personhood).

propose a fictionalizing approach to algorithmic personhood (analogous to the law's fiction of corporate personhood)[83] face two formidable obstacles. First, no one has proposed a satisfactory answer to when it would make sense to hold algorithms responsible. Anything an algorithm does is ultimately a product of its environment and its programming.[84] As such, it is hard to see when the algorithm, rather than its environment or its programmer, would be culpable. For example, in 2016, Microsoft launched a chatbot, "Tay," to communicate with teens online.[85] Tay was supposed to teach itself to talk by learning from data it scraped from Twitter.[86] Within twenty-four hours, internet users had baited Tay with enough corrupting tweets that the bot's messages became chauvinistic, racist, and anti-Semitic.[87] It is far from clear, though, that Tay was to blame for the things it said and not Microsoft or a corrosive "Twitterverse."[88] This leads to the second difficulty with direct algorithmic liability: Even if the law were to find an algorithm like Tay responsible, then what? There is no way to sanction an algorithm or bot[89] (short, perhaps, of killing it, as Microsoft did to Tay).[90] We can jail or fine other "people," but algorithms lack bodies and pocketbooks.[91]

83. *See* Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1239–43 (1992) (comparing the case for AI personhood and corporate personhood).

84. *See* DAVID A. PATTERSON & JOHN L. HENNESSY, COMPUTER ORGANIZATION AND DESIGN: THE HARDWARE/SOFTWARE INTERFACE 13–15 (5th ed. 2014) (ebook); Anupam Chandler, *The Racist Algorithm*, 115 MICH. L. REV. 1023, 1034–37 (2017). This is not to deny that problems with algorithms can arise in other ways, as when, for example, an algorithm designed for one use is put to a different use. *See* Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFO SYS. 330, 330–32 (1996).

85. *See* Elle Hunt, *Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter*, GUARDIAN (Mar. 24, 2016, 2:41 PM), https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter [https://perma.cc/B933-9RQS].

86. *See id.*

87. Damon Beres, *Microsoft Chat Bot Goes on Racist, Genocidal Twitter Rampage*, HUFFPOST (Mar. 24, 2016, 10:19 AM), https://www.huffpost.com/entry/microsoft-tay-racist-tweets_n_56f3e678e4b04c4c37615502 [https://perma.cc/GP9Z-YLG7].

88. *See* James Vincent, *Twitter Taught Microsoft's AI Chatbot To Be a Racist Asshole in Less Than a Day*, VERGE (Mar. 24, 2016, 6:43 AM), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist [https://perma.cc/M4BU-VDVQ].

89. *See* Joanna J. Bryson, Mihailis E. Diamantis & Thomas D. Grant, *Of, For, and By the People: The Legal Lacuna of Synthetic Persons*, 25 ARTIFICIAL INTELLIGENCE & L. 273, 288 (2017); Solum, *supra* note 83, at 1244–48 (discussing difficulties of punishing algorithms). *But see* Gabriel Hallevy, *"I, Robot—I, Criminal"—When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses*, 22 SYRACUSE SCI. & TECH. L. REP. 1, 29–35 (2010) [hereinafter Hallevy, *I, Robot*] (offering unpersuasive accounts of diverse AI punishments).

90. *See* Rob Price, *Microsoft Is Deleting Its AI Chatbot's Incredibly Racist Tweets*, BUS. INSIDER (Mar. 24, 2016, 7:31 AM), https://www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3 [https://perma.cc/8C2R-2LXJ (dark archive)].

91. *See generally* Ryan Abbott & Alex Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, 53 U.C. DAVIS L. REV. 323 (2019) (discussing the difficulties in traditionally "punishing" algorithms for misconduct and ultimately rejecting the possibility of punishing algorithms).

Regardless of whether the law could or should hold algorithms directly liable, its present approach is clear: algorithms are not people and they cannot be civil or criminal defendants.[92] Reversing course would require the swing of a sledgehammer. This Article limits itself to *corporate* liability for algorithmic misconduct because the law has already settled that corporations are responsible "persons." As evidenced by the Federal Sentencing Guidelines provisions on organizations, the law also already has longstanding mechanisms for sanctioning corporations.[93] The surgical approach adopted here draws on that existing legal structure to ask: Under what conditions should *corporations* be liable when their algorithms engage in misconduct? It thereby avoids the conceptual, philosophical, legal, and pragmatic challenges of holding algorithms directly accountable.[94]

Despite the significant difficulties posed by corporate algorithmic misconduct, the legal revisions this Article proposes are relatively modest. The solution it seeks should be an extension, rather than a rewriting, of current law. The solution should embrace the present law of corporate liability, including the fiction of corporate personhood and culpable mental states. It should avoid imposing strict liability on corporations, which is politically unfeasible and would overly impede corporate innovation. Lastly, it should not require a new body of law establishing the legal personhood or accountability of algorithms. Under the solution offered below, algorithms do not think or know things. Rather, corporations think or know things through the algorithms they use.

The general strategy proposed below leverages a quirk of current corporate law according to which corporations are people and can have mental states. The argument steps into that fiction and explores what its implications are. In light of the kind of "minds" the law says corporations have, there is little reason to say that algorithms cannot play a role in determining what corporations think and know.

II. HEALTHCO AND FORMBOT: A CLARIFYING EXAMPLE

The following hypothetical highlights some of the challenges that a solution to the problem of algorithmic corporate misconduct must address:

92. *See* Thomas Beardsworth & Nishant Kumar, *Who To Sue When a Robot Loses Your Fortune*, BLOOMBERG (May 5, 2019, 8:00 PM), https://www.bloomberg.com/news/articles/2019-05-06/who-to-sue-when-a-robot-loses-your-fortune [https://perma.cc/27PX-4E49] ("Robots are getting more humanoid every day, but they still can't be sued.").

93. U.S. SENTENCING GUIDELINES MANUAL § 8A1.1 cmt. n.1 (U.S. SENTENCING COMM'N 2018).

94. This is not to say that the law will not eventually need to find a way to hold algorithms directly accountable. The solution proposed here will not work when algorithms acting alone, and not on behalf of a corporation, engage in misconduct.

> HealthCo is a corporation that provides a wide array of services to Medicare and Medicaid-eligible patients. To save costs, HealthCo asked its data engineers to develop FormBot, a machine-learning algorithm trained to complete and file federal reimbursement forms as efficiently as possible. After several months in operation, FormBot learned on its own that it could secure more reimbursements in less time if it used fake information for some of the forms. Nobody at HealthCo knew about or expected this development. By the time federal authorities discovered the fake forms, HealthCo had received millions of dollars in improper reimbursements.

Did HealthCo violate the criminal[95] or civil[96] provisions of the False Claims Act, which prohibit knowingly submitting false claims to the federal government?

There are several things to note about the HealthCo example. First, it zeroes in on one of the challenges posed by the problem of algorithmic misconduct. There is no doubt under current law that corporations can violate the False Claims Act.[97] HealthCo satisfies the objective elements of the violation because it submitted false claims to the federal government. The sticking point is the False Claims Act's knowledge requirement. If a HealthCo employee had submitted the forms knowing they contained fake information, the case for liability would be pretty clear.[98] However, under the facts of the hypothetical, current law dictates neither civil nor criminal liability. HealthCo could not have known the forms were fake because none of its human employees did. Automatically terminating the liability inquiry in this way is worrying. Public coffers were harmed. The fact that HealthCo used an algorithm rather than an employee does not alter society's interest in deterring or condemning such conduct. At a minimum, the circumstances warrant a more discriminating liability inquiry.

Second, criminal liability is also at issue in the hypothetical. This is important because criminal law arguably has the biggest stake in getting corporate mental states right. Some scholars tout the efficiency benefits of strict liability standards in civil law,[99] particularly in the corporate context.[100] But criminal law concerns itself with more than efficiency.[101] Criminal law would

95. 18 U.S.C. § 287 (2018).

96. 31 U.S.C. § 3729(a)(1)(A)–(G) (2018).

97. *See, e.g.*, Nye & Nissen v. United States, 336 U.S. 613, 614–16 (1949).

98. *See, e.g.*, United States v. Sain, 141 F.3d 463, 470–71 (3d Cir. 1998).

99. *See, e.g.*, Guido Calabresi & Jon T. Hirschoff, *Toward a Test for Strict Liability in Torts*, 81 YALE L.J. 1055, 1060–67 (1972) (advocating for a strict liability test in torts but noting some weaknesses in such an approach).

100. *See* Fischel & Sykes, *supra* note 69, at 327–28.

101. *See* Albert W. Alschuler, *The Changing Purposes of Criminal Punishment: A Retrospective on the Past Century and Some Thoughts About the Next*, 70 U. CHI. L. REV. 1, 1 (2003). *But see* Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 170, 172 (1968) (focusing on an

not fulfill its distinctive condemnatory[102] and powerful deterrent functions[103] if it relied only on strict liability. Strict liability overlooks differences that matter to victims demanding justice and to defendants facing punishment. It treats the innocent dupe, the hapless fool, and the calculating villain all alike.[104] A system of criminal law that makes no reference to mental states would be unrecognizable.[105]

Third, the particular mental state at issue for HealthCo is knowledge.[106] Every type of mental state has its own distinctive properties and would ideally receive a separate discussion. Though there are over one hundred different types of mens rea among the provisions of the federal criminal code alone,[107] knowledge is an element of many of the most common corporate crimes.[108] The solution proposed below, while framed in terms of knowledge, should serve as a template for other mental states. Knowledge is a convenient starting point since that is where existing literature on the extended mind thesis tends to focus.[109]

Lastly, the hypothetical specifies technical details that make the problem of algorithmic misconduct particularly intractable. HealthCo's engineers programmed FormBot using machine learning. Very roughly, machine learning techniques start by specifying the algorithm's goal and then train the algorithm with a large set of test cases.[110] By telling the algorithm in each test case whether

economic and efficiency-based approach to determining criminal punishments). *See generally* Jeremy Bentham, *Principles of Penal Law*, *in* 1 THE WORKS OF JEREMY BENTHAM 366 (John Bowring ed., 1962) (explaining the goals of criminal law and what it offers both to victims and society).

102. Joel Feinberg, *The Expressive Function of Punishment*, 49 MONIST 397, 400 (1965).

103. Samuel W. Buell, *The Blaming Function of Entity Criminal Liability*, 81 IND. L.J. 473, 500, 510–12 (2006).

104. *See generally* Paul H. Robinson & Jane A. Grall, *Element Analysis in Defining Criminal Liability: The Model Penal Code and Beyond*, 35 STAN. L. REV. 681, 687–90 (1983) (discussing how the criminal law uses different mental state elements to determine culpability).

105. This is why Rebecca Crootof, proposing strict state liability for harms caused by warfare AI, turns from criminal law to tort law. Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, 164 U. PA. L. REV. 1347, 1387–88 (2016).

106. It bears noting that the legal definition of knowledge is not the same as the philosophical definition. In the law, a person knows some information if she believes it and it is true. MODEL PENAL CODE § 2.02(2)(b)(i)–(ii) (AM. LAW INST. 1985). Philosophers have additional requirements for knowledge, one of which is that the person also have a justification for her belief. *See* PAUL K. MOSER & ARNOLD VANDER NAT, HUMAN KNOWLEDGE: CLASSICAL AND CONTEMPORARY APPROACHES 3 (2d ed. 1995). I am using "knowledge" in the legal sense consistent with its meaning in the False Claims Act.

107. *See* William S. Laufer, *Culpability and the Sentencing of Corporations*, 71 NEB. L. REV. 1049, 1065 (1992).

108. *See* Diamantis, *Functional Corporate*, *supra* note 76, 322–23 (listing examples).

109. Notable exceptions include Mark Rowlands, *Consciousness, Broadly Construed*, *in* THE EXTENDED MIND 271, 271 (Richard Menary ed., 2010) (ebook), and Mattia Gallotti & Bryce Huebner, *Collective Intentionality and Socially Extended Minds*, 30 PHIL. PSYCHOL. 251, 252–53 (2017).

110. *See* David Lehr & Paul Ohm, *Playing with Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 668 (2017); Jason Brownlee, *Supervised and Unsupervised*

or not it attained its goal, the algorithm can eventually learn to succeed on its own.[111] For example, engineers might want to design a drone-flying algorithm to take the most efficient route to a target.[112] They would code this goal and then train the drone by putting it in various places around the target, seeing where it goes, and telling the algorithm whether its performance was successful or not. If all goes well, the drone should learn to seek out the target reliably and efficiently.

The significance of machine learning for algorithmic misconduct is twofold. First, machine learning is behind the most sophisticated AI systems.[113] As the social threat of corporate algorithmic misconduct expands, machine learning is likely to predominate. Second, machine learning raises the possibility that algorithms will misbehave without any intervening human misconduct.[114] In the hypo, HealthCo's engineers did not design FormBot to submit fake forms, and no one knew or expected that it would. This is realistic because machine learning algorithms effectively write their own code.[115] The resulting algorithms become so complicated that programmers analyzing the code afterwards often cannot understand how it works.[116] Additionally, many algorithms have built-in randomness as an essential part of their design.[117] Consequently, machine learning algorithms can behave in unintended and unanticipated (and unanticipatable) ways.[118] In hindsight, aberrant results often trace back to some feature of the machine learning process: how the goal was specified, the set of test cases used to train the algorithm, or some interaction

Machine Learning Algorithms, MACHINE LEARNING MASTERY (Mar. 16, 2016), https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ [https://perma.cc/DSF4-TZAU].

111. *See A Beginner's Guide to Neural Networks and Deep Learning*, PATHMIND: A.I. WIKI, https://skymind.ai/wiki/neural-network [https://perma.cc/P59K-RJ45].

112. *See* Lemley & Casey, *supra* note 17, at 1313–14 (discussing this example).

113. *See id*. at 1335 ("[T]he unpredictability inherent in machine learning is also one of its greatest strengths.").

114. *See generally* PEDRO DOMINGOS, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD (2015) (outlining how current algorithms function and the positive and potentially negative impacts of machine learning on their functions). Ryan Abbott and Alex Sarch call this behavior "Hard AI Crime[]." Abbott & Sarch, *supra* note 91, at 328.

115. *See* Abbott & Sarch, *supra* note 91, at 330–31.

116. Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1089–90, 1092 (2018); Matthew Carroll, *The Complexities of Governing Machine Learning*, DATANAMI (Apr. 27, 2017), https://www.datanami.com/2017/04/27/complexities-governing-machine-learning/ [https://perma.cc/H3EU-LB4J].

117. Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 655 (2017).

118. Lemley & Casey, *supra* note 17, at 1365 ("[M]uch of the [algorithmic] misconduct that tomorrow's designers, policymakers, and watchdogs must guard against might not be intentional at all.").

between these two factors and the real world.[119] Due to the code's complexity, problems can arise even if every human involved is fully innocent.[120]

In the drone example from above, the engineers at one point observed the drone doing the exact opposite of what they had intended—flying away from the target to the perimeter of the test area.[121] After some investigation, they found that this was not a malfunction. The drone learned that in some circumstances, the most efficient way to get to the target was to fly to the perimeter of the test area.[122] If the drone did that, it learned that the engineers would retrieve it and carry it to the target to reset the trial. Being carried rather than flying was sometimes the most efficient route to the target, though that is clearly not what the engineers intended the drone to do.

The fact that FormBot used machine learning and filed fake forms without anyone designing it to do so ensures that the violation in the hypothetical is a pure case of algorithmic misconduct.[123] Otherwise, the case might just involve ordinary employee misconduct, albeit misconduct mediated by an algorithm. The law already has mechanisms to handle cases where employees purposely, knowingly, or recklessly design algorithms to break the law.[124] In the hypothetical, if a HealthCo engineer purposely or knowingly designed FormBot to submit fake forms, respondeat superior would attribute the engineer's mental state to HealthCo, thereby satisfying the False Claims Act's requirement.[125] It would not matter that the engineer did not physically submit the forms herself.[126] If the engineer recklessly designed FormBot, willful

119. *See, e.g.*, Kroll et al., *supra* note 117, at 693–94; Lemley & Casey, *supra* note 118, at 1313.

120. KEVIN PETRASIC ET AL., ALGORITHMS AND BIAS: WHAT LENDERS NEED TO KNOW 1 (2017), https://www.whitecase.com/sites/whitecase/files/files/download/publications/algorithm-risk-thought-leadership.pdf [https://perma.cc/36YD-VWTW] ("[A] perfectly well-intentioned algorithm may inadvertently generate biased conclusions that discriminate against protected classes of people."); Barocas & Selbst, *supra* note 29, at 729 (explaining that errors "may be the result of entirely innocent choices made by data miners").

121. Lemley & Casey, *supra* note 17, at 1313.

122. *Id.*

123. It is thus unlike some examples that other criminal scholarship has focused on. *See, e.g.*, Amanda McAllister, Note, *Stranger than Science Fiction: The Rise of A.I. Interrogation in the Dawn of Autonomous Robots and the Need for an Additional Protocol to the U.N. Convention Against Torture*, 101 MINN. L. REV. 2527, 2545, 2547 (2017).

124. *See* Hallevy, *I, Robot*, *supra* note 89, at 9 (discussing different models of criminal liability for crimes involving robots premised on programmer or user fault).

125. *See* MODEL PENAL CODE § 2.02(5) (AM. LAW INST. 1985).

126. Causing a claim to be submitted falsely explicitly satisfies the civil version of the False Claims Act. 31 U.S.C. § 3729(a)(1)(A) (2018). For criminal law, other liability doctrines fill the gap. *See* 18 U.S.C. § 2(b) (2018) ("Whoever willfully causes an act to be done which if directly performed by him or another would be an offense against the United States, is punishable as a principal."); MODEL PENAL CODE § 5.04(1)(a) (AM. LAW. INST. 1985) (stating that a person may be guilty of soliciting an innocent functionary to commit a crime). The actus reus and mens rea elements of a crime may be satisfied by different parts of the corporation. *See, e.g.*, United States v. Bank of New England, N.A., 821 F.2d 844, 856 (1st Cir. 1987).

ignorance might provide a basis for saying she, and hence HealthCo, had constructive knowledge of the fake forms.[127] What current doctrine cannot handle, and what this Article addresses, are the most worrying cases where employee misconduct is removed from the picture.[128] A solution for cases of purely algorithmic corporate misconduct will also provide an alternate route to corporate liability in situations where an individual employee may have been at fault, but where proving so is difficult.[129]

III. THE EXTENDED MIND THESIS

It is time to modernize the law's conception of corporate mentality. The law can only properly hold corporations accountable by creating the possibility that they sometimes "know" the information readily available on their servers and "intend" the decisions reached by their algorithms. Many cognitive scientists and philosophers understand mental states in a way that could lay the foundation for reaching beyond respondeat superior's exclusive focus on employees. The present part describes this understanding as applied to natural people. The next part broadens the theory and shows how to adapt it, both in principle and in law, to corporate people.

The "extended mind thesis" states that the human mind is not always constrained by the physical boundaries of the brain.[130] Extended mind theorists typically endorse a "functionalist" account of mental states.[131] According to functionalism, mental states are characterized by the cognitive role they play connecting inputs (like environmental cues and other mental states) to outputs (like new mental states or behavior).[132] For example, if a person desires ice cream and walks to the freezer, there is a good chance it is because she believes ice cream is there. A rough defining characteristic of belief (e.g., that there is ice cream in the freezer) is that it relates desire inputs (e.g., for ice cream) and

127. *See* Glob.-Tech Appliances, Inc. v. SEB S.A., 563 U.S. 754, 769 (2011) (explaining that, in the Court's view, willful blindness "surpasses recklessness and negligence").

128. Or not provable in the picture. *See generally* Memorandum from Eric Holder, Deputy Attorney Gen., to All Component Heads & U.S. Attorneys (June 16, 1999), http://www.justice.gov/sites/default/files/criminal-fraud/legacy/2010/04/11/charging-corps.PDF [https://perma.cc/6BH5-D2FC] (discussing the procedure for bringing criminal charges against corporations and the difficulties in identifying wrongdoers).

129. *See* Barocas & Selbst, *supra* note 29, at 692–93 (describing how computer programs can mask human misconduct); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 884–85 (2017) (discussing misconduct and algorithms in the context of "intentional discrimination").

130. Clark & Chalmers, *supra* note 43, at 14.

131. *See, e.g.*, *id.* ("What makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body."); Richard Menary, *Introduction: The Extended Mind in Focus to* THE EXTENDED MIND, *supra* note 109, at 1, 5 (describing the "functionalist credentials of [extended mind theory]"); Michael Wheeler, *In Defense of Extended Functionalism*, *in* THE EXTENDED MIND, *supra* note 109, at 245, 245.

132. GILBERT HARMAN, REASONING, MEANING, AND MIND 236–39 (1999) (ebook).

behavioral outputs (e.g., walking to the freezer).[133] According to functionalism, any state that appropriately relates desires and behavior in this way could be a belief.

One important corollary of functionalism is its neutrality about the systems that realize mental states. For example, it does not matter what the system is made of.[134] Human neurons clearly can do the job. But so could systems made of complex arrangements of different material, whether organic (e.g., animal or alien brains) or inorganic (e.g., very sophisticated arrangements of cogs or circuits).

Similarly, functionalism also does not care where the systems realizing mental states reside. The seminal work in extended mind theory puts the point as a "Parity Premise": "If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* . . . part of the cognitive process."[135] Typically, the systems underwriting human mental states are located within the skull. But a person whose brain protrudes beyond her skull could still have mental states in the protruding part. This could be true even if her brain part was quite distal, connected by long neurons. All that would matter is that the part had everything it needed (the right internal organization and the right connections to the rest of the brain) to carry out mental state functions.

Combining the insights of the previous two paragraphs—neutrality as to material and as to location—allows extended mind theorists to talk about a broad range of more meaningful examples. The following two cases illustrate[136]:

> Alice wants to walk from her house to a new café. She looks up the directions on her computer and commits them to memory. She then sets off and easily finds her way.
>
> Barry also wants to walk from his house to the new café. He suffers from Alzheimer's and has trouble remembering things. He looks up the directions on his computer and carefully writes them into his diary. He then sets off. By checking his diary for guidance at each turn, he easily finds his way.

133. *See* BRIAN LOAR, MIND AND MEANING 6–9 (1981) (explaining the belief-desire theory).

134. *See* Menary, *supra* note 131, at 6 ("[I]t is not the physical properties that matter to [extended mind theory], however, but the functionality of the process."); Hilary Putnam, *Psychological Predicates*, *in* ART, MIND, AND RELIGION 37, 44–45 (W.H. Capitan & D.D. Merrill eds., 1967); J.J.C. Smart, *Sensations and Brain Processes*, 68 PHIL. REV. 141, 150 (1959).

135. Clark & Chalmers, *supra* note 43, at 8; Menary, *supra* note 131, at 5.

136. The following hypothetical is based on the hypothetical used by Andy Clark & David Chalmers. *See id.* at 12–16 (discussing a similar hypothetical in which Inga uses her memory to get to a museum and Otto uses a notebook).

There is no doubt that Alice knew how to get to the store after she looked up the directions. But did Barry? His diary entries seem to play a functional role similar to the directional information encoded in Alice's neurons. Input: desire to visit café. Output: accurately and easily walking there. Alice and Barry share the same functional relationship between inputs and outputs. Extended mind theorists conclude that since Alice knew how to get to the café, so did Barry.

To resist that conclusion, critics of the extended mind thesis need to find a meaningful difference between Alice and Barry. As it happens, that is difficult to do. One obvious difference is that Barry's "knowledge" is outside his cranium and encoded on paper. Alice's is in her skull and encoded in neurons. But the significance of that difference is precisely what the extended mind thesis calls into doubt. Relying on it to defeat the thesis would be question-begging. Another possible difference is that Alice may have faster recall of the information, while Barry has to take time to look it up in his diary. But that difference, while not question-begging, is also not meaningful. To see why, consider what it would mean for Alice. What if she had to ponder at each turn (for longer than it took Barry to read his diary) before recollecting which way to go? Clearly that would not undermine her claim to know the directions. Maybe the relevant difference between Alice and Barry is that Barry could lose his diary on the way.[137] But Alice could also lose her memories: a falling tree branch or a stroke could disrupt her fragile neural connections. In the absence of a meaningful difference between Alice and Barry, advocates of the extended mind thesis reaffirm that Barry must have known the directions. His mind extends to the diary pages on which the directions are written.

Moving beyond ad hoc arguments about individual cases, extended mind theorists propose general criteria for evaluating when a person counts as knowing externally housed information. These criteria are supposed to capture the functional relationship between a person and information she knows in paradigmatic cases. Assuming a person uses information to direct her behavior,[138] the most commonly accepted criteria are that

1. the information is available and the subject typically invokes it;
2. the subject more or less automatically endorses (i.e., is prepared to act on and reason with) the information upon retrieval; and

137. *See id.* at 15 (discussing the fact that, in a similar hypo, a notebook on which memories are written or stored may be taken away). *See generally* Fred Adams & Kenneth Aizawa, *The Bounds of Cognition*, 14 PHIL. PSYCHOL. 43, 55–56, 62–63 (2001) (noting the differences in the cognitive process between memorizing directions and following written directions); Robert D. Rupert, *Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind?*, *in* THE EXTENDED MIND, *supra* note 109, at 325, 325 (arguing that "external bits of language do not become part of [the cognitive] system").

138. Extended mind theorists call this "causal coupling" and set out criteria for it. Menary, *supra* note 131, at 3–4.

3. the subject can easily access the information.[139]

Both Alice and Barry satisfy these conditions with respect to their directional information. More generally, anyone who uses information to direct her behavior and satisfies the conditions counts as knowing it, regardless of where or how the information is stored: on neurons, diaries, tattoos, rolodexes, cell phones, laptops, wherever.

While the last few paragraphs focused on philosophical arguments, many cognitive scientists also endorse the extended mind thesis. They recognize several types of extended cognitive systems.[140] Some of these systems involve an individual person using external objects. For example, a person may use fingers or pebbles as an aid to long-form arithmetic when juggling several numbers in memory proves difficult.[141] Other extended cognitive systems are made up of multiple individuals, e.g., navigational teams,[142] large-scale scientific research,[143] and transactive memory systems.[144] Cognitive scientists draw on themes reflected in the extended mind thesis to explain phenomena in situated cognition,[145] robotics,[146] and child development.[147] The theory of embodied cognition provides the framework for this perspective: "Many features of cognition . . . are deeply dependent upon characteristics of the physical body

139. Andy Clark, *Mementos Revenge: The Extended Mind, Extended*, *in* THE EXTENDED MIND, *supra* note 109, at 43, 46.

140. I should also note that this view is far from uncontroversial in cognitive science. *See, e.g.*, ROBERT D. RUPERT, COGNITIVE SYSTEMS AND THE EXTENDED MIND 61 (2009) (critiquing a realization-based argument for extended cognition); Fred Adams & Kenneth Aizawa, *Why the Mind is Still in the Head*, *in* THE CAMBRIDGE HANDBOOK OF SITUATED COGNITION 78, 78 (Philip Robbins & Murat Aydede eds., 2009) (arguing against the extended mind theory and stating that "the mind is still in the head").

141. *See generally* MERLIN DONALD, ORIGINS OF THE MODERN MIND (1991) (exploring the emergence of visual symbolism and external memory as a major evolutionary transition for humans' cognitive ability).

142. *See* EDWIN HUTCHINS, COGNITION IN THE WILD 26 (1995) (discussing navigational teams); *see also* Edwin Hutchins, *The Social Organization of Distributed Cognition*, *in* PERSPECTIVES ON SOCIALLY SHARED COGNITION 283, 305–06 (Lauren B. Resnik et al., eds., 1991) (discussing group cognitive networks and properties).

143. *See* Ronald N. Giere & Barton Moffatt, *Distributed Cognition: Where the Cognitive and the Social Merge*, 33 SOC. STUD. SCI. 301, 301–03 (2003).

144. Transactive memory is a system in which a group of people encodes, stores, and retrieves data. *See* Daniel M. Wegner, *A Computer Network Model of Human Transactive Memory*, 13 SOC. COGNITION 319, 319–20 (1995).

145. *See generally* LUCY A. SUCHMAN, PLANS AND SITUATED ACTIONS: THE PROBLEMS OF HUMAN MACHINE COMMUNICATION (1987) (exploring how human action is related to social and physical circumstances).

146. *See generally* RANDALL D. BEER, INTELLIGENCE AS ADAPTIVE BEHAVIOR: AN EXPERIMENT IN COMPUTATIONAL NEUROETHOLOGY 1–18 (B. Chandrasekaran ed., 1990) (discussing different theories of intelligence and the idea of adaptive intelligence).

147. *See generally* ESTHER THELEN & LINDA B. SMITH, A DYNAMIC SYSTEMS APPROACH TO THE DEVELOPMENT OF COGNITION AND ACTION (1996) (ebook) (discussing generally the development of cognitive abilities in young children).

. . . [which] play[] a significant causal role . . . in that agent's cognitive processing."[148] The basic idea is that minds are not so much tools for "thinking," but tools for doing things in the world.[149] From there, it is a short step to the extended mind thesis.[150] Conceptualizing the mind as a tool naturally lends itself to a functional characterization linking environmental cues and accomplished tasks. The systems that help us accomplish those tasks in the presence of the specified environmental cues qualify as part of our minds.

IV. EXTENDING THE CORPORATE MIND

The extended mind thesis suggests that the traditional boundaries defining where mental states reside are too restrictive. In the context of individual humans, the thesis means that the mind is not limited to the brain: it can extend to external cognitive aids like diaries and cell phones. If the thesis carries over to the corporate context, then corporate minds can also extend beyond their traditional limits—the minds of individual employees—to include other functionally integrated corporate systems.

Though extended mind advocates have so far only talked about natural people,[151] parallel arguments could apply for corporate people too. The starting premise is that corporations have minds. The law takes care of that premise by directing us to assume corporations are people.[152] Corporate mental states, the law presently tells us, reside within the heads of employees. A functionalist understanding of mental states implies that any systems carrying out the same functional roles as employees could also form part of the corporate mind.[153] The Parity Premise, quoted in the previous section, could easily adapt to the

148. Robert A. Wilson & Lucia Foglia, *Embodied Cognition*, STAN. ENCYCLOPEDIA PHIL. (Dec. 8, 2015), https://plato.stanford.edu/entries/embodied-cognition/ [https://perma.cc/X9F5-QG5E].

149. *See* ANDY CLARK, BEING THERE: PUTTING BRAIN, BODY, AND WORLD TOGETHER AGAIN 196 (1997) (ebook).

150. *See generally* Rupert, *supra* note 137 (discussing cognitive processing and the extended mind thesis).

151. One fascinating article argues that group minds could be formed from the extension of individual minds to other individual minds. Deborah Perron Tollefsen, *From Extended Mind to Collective Mind*, 7 COGNITIVE SYS. RES., 2006, at 140, 140–41 (2006). In Tollefsen's view, the group mind is the result of the extension, not (as I propose here) the mind which is extended. *See id.* at 146. She explicitly excludes AI from her view. *See id.* at 141 (stating that her article focused on collective systems "constituted primarily by humans").

152. *See* United States v. A&P Trucking Co., 358 U.S. 121, 123, 125 (1958) ("[I]t is elementary that such impersonal entities can be guilty of 'knowing' or 'willful' violations of regulatory statutes through the doctrine of *respondeat superior*.").

153. The National Highway Traffic Safety Administration seems open to a similar sort of functional reasoning. *See* Nat'l Highway Traffic Safety Admin., Opinion Letter on Applicability of Federal Motor Vehicle Safety Standards to Google's Self-Driving Vehicles (Feb. 4, 2016), https://isearch.nhtsa.gov/files/Google%20--%20compiled%20response%20to%2012%20Nov%20%2015%20interp%20request%20--%204%20Feb%2016%20final.htm [https://perma.cc/7MQF-YD95] ("If no human occupant of the vehicle can actually drive the vehicle, it is more reasonable to identify the 'driver' as whatever (as opposed to whoever) *is* doing the driving.").

corporate context: if a part of the world functions as a corporate process which, were it performed by an employee, the law would accept as part of the corporate mind, then that part of the world is part of the corporate mind.[154] Significantly, that "part of the world" could be a smart algorithm running corporate operations. This extension of the corporate mind from human employees to automated algorithms would strike many computer scientists as quite natural. The classic definition of artificial intelligence is functional: any algorithm that "mak[es] a machine behave in ways that would be called intelligent if a human were so behaving."[155]

The argument for the extended mind thesis should be even easier for corporations than it is for humans. Corporate minds are, like corporations themselves, socially constructed objects.[156] It is solely by dint of legal fiat[157] that corporations have minds at all or that they share the mental states of their employees.[158] This removes the main intuitive barriers to the extended mind thesis. For human beings, there is a natural alternative to saying the mind encompasses all functionally integrated cognitive aides. That alternative is grounded in common-sense biology—the mind is limited by the brain.[159] But a plausible limiting principle is much harder to come by where the corporate mind is concerned.[160] Corporations have no brains of their own.[161]

The remainder of this part discusses how to integrate the extended mind thesis into the law of corporate liability. Following two accounts of what the

154. *See* Clark & Chalmers, *supra* note 43, at 8.

155. John McCarthy et al., Proposal for the Dartmouth Summer Research Project on Artificial Intelligence 7 (Aug. 31, 1955) (unpublished manuscript), http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf [https://perma.cc/H9L6-S8D7]. *See generally* A.M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433 (1950) (discussing broadly whether machines can think).

156. *See* Trs. of Dartmouth Coll. v. Woodward, 17 U.S. (4 Wheat.) 518, 636 (1819) ("A corporation is an artificial being, invisible, intangible, and existing only in contemplation of law.").

157. And perhaps some psychological projection on our part. *See* Diamantis, *Corporate Criminal*, *supra* note 35, at 2077–80.

158. The various alternative approaches laws take to define the corporate mind reflect the metaphysical arbitrariness of respondeat superior. *See Crim. Code Act of 1995* (Cth) div 12.3(2)(c) (Austl.) (setting out the corporate ethos approach); United States v. Bank of New England, N.A., 821 F.2d 844, 853, 856 (1st Cir. 1987) (setting out the collective knowledge doctrine); MODEL PENAL CODE § 2.07(1)(c) (AM. LAW INST., Proposed Official Draft 1962) (explaining that a corporation can be convicted based on decisions or actions of the board of directors or a high managerial agent acting within the scope of their employment).

159. A sophisticated version of this challenge is the so-called "coupling-constitution fallacy." *See* FREDERICK ADAMS & KENNETH AIZAWA, THE BOUNDS OF COGNITION 76–105 (2010) (ebook).

160. For similar reasons, the objection to the extended mind thesis that people lack direct access to information in extended systems, *see generally* John Preston, *The Extended Mind, The Concept of Belief, and Epistemic Credit*, *in* THE EXTENDED MIND, *supra* note 109, at 355 (discussing access, authority, and belief in the context of the extended mind theory), does not apply to corporations.

161. *See* United States v. Ladish Malting Co., 135 F.3d 484, 492 (7th Cir 1998) ("Corporations do not record knowledge in neural pathways; they record it in file cabinets (and increasingly on computer disks).").

proposed doctrine might look like, this part takes up policy considerations that could call for pragmatic limits on how far the corporate mind extends.

A. *Doctrinal Proposal*

Under what conditions does a corporation know information embedded in algorithms or big data systems? According to centuries-old legal doctrine, corporations know things stored in employee brains. Extended mind theory offers a framework for reaching beyond that traditional perimeter to include digitally stored information. As was the case when discussing human minds, there are two approaches for evaluating particular cases: coming up with ad hoc analogies and applying generalized criteria. Both approaches should produce largely the same results, though one or the other may be more useful depending on context.

1. The Analogical Approach

The previous part's discussion of Alice and Barry illustrates the analogical approach.[162] Stated abstractly, the approach compares a person of interest, P_1, who bears a functional relationship to some information, I_1, with a second person, P_2, who has a relevantly similar functional relationship to some similar information, I_2. If it is clear that P_2 knows I_2, then P_1 must know I_1.[163] This is the upshot of defining mental states in terms of their functional role—same functional relationship, same mental state.

Carrying out the analogical approach for corporate people works slightly differently from how it works for natural people. Where P_1 is a natural person (like Barry), the readiest comparator P_2 with clear knowledge is someone who (like Alice) had the information stored in her brain. However, where P_1 is a corporate person with digitally-stored I_1, the analogical approach calls for a different sort of P_2 comparator. Under respondeat superior, the clear-cut case of corporate knowledge is one where the information is stored in an *employee* brain. So P_2 should be a corporation that behaves similarly to P_1, but where it is clear that an employee knows I_2. The success of the analogy turns on two factors: (1) whether the functional relationships between P_1 and I_1 and P_2 and I_2 are relevantly similar and (2) how obvious it is that P_2 (i.e., an employee at P_2) knew I_2.[164]

162. *See supra* Part III.

163. Some scholars discussing negligence have proposed different standards for employees and AI, e.g., that the standards for negligence in AI should be twice as strict as those for human individuals. *See* Geistfeld, *supra* note 17, at 1679 (reasoning that automated cars should be "at least twice as safe" as man operated vehicles before being put on the road).

164. The analysis here presumes that we have a workable theory of when AI behavior is attributable to corporations. So far as I know, we do not. There are several possibilities. Attribution might turn on whether the corporation owns the relevant software. Or whether the corporation owns the hardware running the software. Or whether the corporation subsequently endorsed the behavior. While I would

The analogical approach could be easily adapted to the fact-finding process at trial. Hypothetical and comparative reasoning are already essential features of the adjudicative process.[165] Indeed, other scholars have emphasized the importance of comparative reasoning for evaluating corporate mental states.[166] In such reasoning, the task for the plaintiff or prosecution is to construct the comparison case. The task for the defense is to question their similarity. And the task of the fact finder is to arbitrate the persuasiveness of the comparison.

It bears emphasizing that civil plaintiffs and criminal prosecutors using the analogical approach would still need to satisfy their relevant burdens of proof: preponderance of the evidence[167] and beyond a reasonable doubt,[168] respectively. It would not be enough for them simply to stipulate in comparison corporation, P_2, that an employee knew I_2. That would just demonstrate that there is *some* comparable P_2 who knows I_2. The strongest implication fact finders could draw from such a case is the mere *possibility* that the corporate defendant, P_1, knew I_1. Strategically, plaintiffs and prosecutors should instead present a P_2 that had access to the relevant information, behaved similarly to P_1, and did so only using employees. Plaintiffs and prosecutors should not stipulate that one of P_2's employees actually knew the information. To satisfy the burden of proof for liability, the fact finders would have to infer that P_2 (i.e., an employee within P_2) most likely knew (civil law) or must have known (criminal law) the relevant information. That would show not just that there is *some* analogous P_2 that knew the information but that *any* analogous P_2 probably did. The implication that would follow is that P_1 probably did too.

When laid out formally, this style of reasoning may seem complex, but it is actually an intuitive process that judges and juries use all the time. It involves evaluating the likelihood that some fact remains true in a hypothetical case with facts similar to the actual case. In civil law, res ipsa loquitur arguments—i.e., that the sort of accident at issue does not ordinarily occur without negligence[169]—have an identical logical structure. The plaintiff effectively argues that in any relevantly similar case, the person who caused the accident

tentatively propose that the relevant variable should be the level of control the corporation exercises over the AI, readers should proceed with their own preferred theory of corporate behavior.

165. *See generally* California v. Carney, 471 U.S. 386, 389, 393 (1985) (using comparative reasoning); Scott Brewer, *Exemplary Reasoning: Semantics, Pragmatics, and the Rational Force of Legal Argument by Analogy*, 109 HARV. L. REV. 923, 925–26 (1996) (explaining that analogy is a frequently used technique in legal arguments).

166. *See, e.g.*, William S. Laufer, *Corporate Bodies and Guilty Minds*, 43 EMORY L.J. 647, 701, 704 (1994) [hereinafter Laufer, *Corporate Bodies*].

167. *See* Herman & MacLean v. Huddleston, 459 U.S. 375, 387 (1983) ("In a typical suit for money damages, plaintiffs must prove their case by a preponderance of the evidence.").

168. *See In re* Winship, 397 U.S. 358, 361 (1970) (discussing the requisite level of proof in criminal cases).

169. *Res Ipsa Loquitur*, BLACK'S LAW DICTIONARY, *supra* note 61.

most likely behaved negligently.[170] Similar reasoning appears in corporate criminal law. Corporate prosecutors can show that, given the circumstances of the misconduct, some employee must have possessed culpable knowledge, even if the prosecutor cannot show who.[171] The structure of that argument requires demonstrating that in any relevantly similar case, an employee must have possessed culpable knowledge.

Consider how the analogical approach would play out in the HealthCo hypothetical. To argue that HealthCo knew the forms were fake, the prosecution would need to present a hypothetical comparison case with three crucial features:

1. The corporation behaved similarly (i.e., filed fake reimbursement forms); and
2. the corporation did so using employees (i.e., rather than an algorithm like FormBot);
3. who had access to (without stipulating that they knew) similar information (i.e., the information that should have been on the forms).

The strength of the prosecution's argument would turn on the credibility of the comparison case, the closeness of its functional similarity to the corporate defendant, and how likely it is that some employee in the comparison case must have known the information.

Like so much in the fact-finding process, the analysis in most cases will not be clear-cut. More facts would be needed in the HealthCo hypothetical before a resolution started to crystalize. It may turn out, for example, that the misstatements on HealthCo's forms were all relatively small. In the hypothetical comparison cases where an employee filled out the forms, this could be consistent with the employee making rounding errors rather than filling out forms she knew to be false. However, it may also turn out that the rounding errors always favored HealthCo. That programmatic behavior, carried out by an employee, would be more consistent with knowing falsification.

2. Using Generalized Criteria

In some situations, it may be helpful to have generalized criteria for evaluating whether a corporation knows information embedded in its algorithms. The criteria would need to characterize the functional relationship

170. Another civil example comes from product liability, where factfinders can infer a product defect from product performance that "was of a kind that ordinarily occurs as a result of product defect." RESTATEMENT (THIRD) OF TORTS: PRODS. LIABILITY § 3(a) (AM. LAW INST. 1997).

171. *See Developments in the Law—Corporate Crime: Regulating Corporate Behavior Through Criminal Sanctions*, 92 HARV. L. REV. 1227, 1248 (1979) (citing United States v. T.I.M.E.-D.C., Inc., 381 F. Supp. 730, 735 n.7, 739 (W.D. Va. 1974)).

corporations bear to information that current law clearly treats them as knowing. Respondeat superior defines what those situations are. If the criteria accurately capture that functional relationship, functionalism dictates that any corporation satisfying the criteria for some information would count as knowing that information. This would be true regardless of how or where the information was stored.

The basic requirement of respondeat superior is that some employee knows the information.[172] Since this first requirement ultimately concerns knowledge of natural people (the employees), slightly modified versions of the criteria extended mind theorists already offer for humans should work:

1. The information is available and the employee/algorithm (on behalf of the corporation) typically invokes it;
2. the employee/algorithm (on behalf of the corporation) more or less automatically endorses the information upon retrieval; and
3. the employee/algorithm (on behalf of the corporation) can easily access the information.

Collectively, these criteria intuitively capture the functional relationship a corporation has with corporate information stored in its employees' brains—the employees have easy recall of that information and use it to perform their job. By replacing "employee" with "algorithm," the criteria easily adapt to answer whether a corporation knows information through one of its algorithms. Most corporate algorithms that have ready access to information they use to direct corporate operations would satisfy the criteria.

Respondeat superior has two additional requirements for attributing employee knowledge to corporations: first, that the employee knows the information within the scope of her employment and, second, that she uses it with an intent to benefit her employer.[173] These two requirements, however, have been so weakened by the courts that it is questionable whether the generalized criteria developed here for knowledge need to account for them. An employee effectively counts as working within the scope of her employment whenever she is on the job, even if she is acting contrary to her employer's orders.[174] And an employee counts as intending to benefit her employer even when that intent is subsidiary,[175] hypothetical,[176] or ineffective.[177]

172. *See supra* Part I.

173. *See* 18B AM. JUR. 2D *Corporations* §§ 1812, 1817, Westlaw (database updated Feb. 2020).

174. *See, e.g.*, United States v. Hilton Hotels Corp., 467 F.2d 1000, 1004 (9th Cir. 1972).

175. *See* United States v. Automated Med. Labs., Inc., 770 F.2d 399, 407 (4th Cir. 1985).

176. *See* United States v. Sun-Diamond Growers of Cal., 138 F.3d 961, 970 (D.C. Cir. 1998), *aff'd*, 526 U.S. 398 (1999).

177. *See* Old Monastery Co. v. United States, 147 F.2d 905, 908 (4th Cir. 1945).

The intuitive purpose of respondeat superior's scope-of-employment and intent-to-benefit requirements is to let corporations off the hook in two scenarios: where an employee knows something only in her private capacity ("My dad didn't really have that surgery.") or goes rogue and uses her knowledge only to thwart corporate goals ("This is how I could embezzle corporate assets."). Since algorithms do not have private lives, the first scenario is irrelevant. However, algorithms can advance or hinder corporate goals. A broken algorithm can victimize a corporation as much as rogue employees can.[178] The following fourth criterion should accommodate the interests behind respondeat superior's intent-to-benefit requirement:

4. Furthermore, the algorithm must use the information in a way that accrues some (perhaps illegitimate or minor) benefit to the corporation.

Like the law's current intent-to-benefit requirement, the overwhelming majority of cases will easily satisfy this fourth criterion.

How would the criteria apply to HealthCo? HealthCo's form filing algorithm, FormBot, clearly satisfies the fourth criterion (benefit to the corporation) since the falsified forms generated millions of dollars for HealthCo. To evaluate the other criteria, more technical details about how FormBot accesses and uses information would be needed. It is probably safe to assume that FormBot satisfies the first (typically invokes the information) and third (automatically endorses the information) criteria. There is no reason HealthCo's engineers would design FormBot with obstructed access to claims information or so that Formbot would not typically invoke that information when filing claims. With respect to the second criterion (automatic endorsement of information), there is more wiggle room. The engineers might have designed FormBot to accept and use the claims information it received uncritically. This would strengthen the claim that FormBot satisfies the second criterion and, consequently, that HealthCo knew the information. Alternatively, HealthCo's engineers might have designed FormBot to be more skeptical, with built-in audit controls to verify and validate the information before using it. In that case, FormBot would be less likely to satisfy the second criterion, and HealthCo would be less likely to count as knowing the information.

178. *See Sun-Diamond*, 138 F.3d at 970 (noting that the employee intended to defraud his employer rather than benefit him). When an algorithm improperly discriminates against loan or job applicants, the company on whose behalf it is working loses out too because they miss out on profitable lending opportunities or quality candidates. *See* Margareta Drzeniek-Hanouz, *Why Discrimination Is Bad for Business*, WORLD ECON. F. (Mar. 6, 2015), https://www.weforum.org/agenda/2015/03/why-discrimination-hurts-competitiveness/ [https://perma.cc/H7R9-BH6F].

It is worth pausing briefly to note how the four criteria generate attractive policy results. They serve as a workable and theoretically grounded basis for bringing extended mind theory into corporate law. As such, they offer the possibility of holding corporations like HealthCo liable when their algorithms break the law. The criteria do not automatically impose liability every time a corporate algorithm causes harm. They have a contingency that gives corporations socially beneficial incentives. For example, from the ex ante perspective, the criteria would have allowed HealthCo to reduce its prospect of liability by building additional quality controls into FormBot. Incentivizing responsible algorithm development is exactly what the law should be doing. It is to this and other policy considerations that the Article now turns.

B. *Policy-Based Objections and Restrictions*

Extending the corporate mind using either of the two approaches just described would be a good first stab at solving the problem of algorithmic corporate misconduct. Though the proposals are a sure improvement over current law—which effectively shields corporations from liability for many algorithmic harms—further refinements might advance corporate law's goals even better. The discussion that follows focuses on criminal justice policies; related considerations arise in civil contexts too.

The foremost policy goals in corporate criminal law are familiar from criminal law more broadly,[179] namely retribution[180] and deterrence.[181] Retribution may initially seem an odd fit for corporate criminal law since corporations are not ordinary moral agents.[182] However, retributive sentiments are a strong driver behind corporate criminal law.[183] There are different versions of retribution theory. The version that best fits corporate criminal law seeks to use criminal liability to vindicate the public's intuitions about when corporations deserve moral condemnation.[184] Even though corporations may not be true moral agents, they occupy a space in our sociopsychology that makes moral judgments about them natural and irresistible.[185] Illustrative is the clarion

179. *See* Meir Dan-Cohen, *Sanctioning Corporations*, 19 J.L. & POL'Y 15, 22 (2010) (listing policy goals).

180. Regina A. Robson, *Crime and Punishment: Rehabilitating Retribution as a Justification for Organizational Criminal Liability*, 47 AM. BUS. L.J. 109, 110 (2010).

181. Darryl K. Brown, *Street Crime, Corporate Crime, and the Contingency of Criminal Liability*, 149 U. PA. L. REV. 1295, 1325 (2001).

182. Alschuler, *Two Ways To Think*, *supra* note 68, at 1392 (highlighting the peculiarity of punishing a corporation or imaginary person as if they were a real individual); Baker, *supra* note 67, at 350.

183. *See* Baer, *supra* 58, at 621 (discussing retributivist sentiments in corporate criminal law).

184. *See* PAUL ROBINSON, INTUITIONS OF JUSTICE AND THE UTILITY OF DESERT 176–88 (2013); Dan M. Kahan & Martha C. Nussbaum, *Two Conceptions of Emotion in Criminal Law*, 96 COLUM. L. REV. 269, 352 (1996).

185. *See* Diamantis, *Corporate Criminal*, *supra* note 35, at 2077–80.

call for justice against corporations that have pushed addictive opioids onto desperate consumers[186] or have destroyed delicate environmental habitats.[187] The vigor of the call warns the criminal justice system not to turn a deaf ear.

According to deterrence theory, criminal liability should seek to prevent corporate misconduct by raising the costs of violating the law.[188] The law can do this by using the threat of sanctions to induce corporations to run their businesses more carefully. Ordinarily, this means corporations will implement additional compliance programs for things like employee training and monitoring.[189] Where algorithms are concerned, taking care means designing algorithms that are less likely to break the law. While nothing can guarantee that a machine learning algorithm will always follow the law (nor can anything guarantee employees will always follow the law either),[190] software engineers can take steps to reduce the probability that algorithms will misbehave.[191] These steps include: diversifying the body of engineers writing algorithms,[192] more careful initial programming,[193] more mindful selection of training data sets,[194]

186. *See, e.g.*, Maia Szalavitz, *Big Pharma's Opioid Greed Was Even Worse than We Thought*, VICE (Sept. 13, 2018, 3:41 PM), https://www.vice.com/en_us/article/7xj97q/big-pharmas-opioid-greed-was-even-worse-than-we-thought [https://perma.cc/2ZKH-VSWL] ("Like the apocryphal child who murdered his parents and then pleaded for sympathy because he'd become an orphan, Purdue first profitably pushed an addictive drug, and then apparently sought to make even more money by treating addictions it helped cause.").

187. *See, e.g.*, George Monbiot, *Shell Is Not a Green Savior. It's a Planetary Death Machine*, GUARDIAN (June 26, 2019), https://www.theguardian.com/commentisfree/2019/jun/26/shell-not-green-saviour-death-machine-greenwash-oil-gas [https://perma.cc/JP2Y-XPHT] ("Trumpeting its investment in natural ecosystems looks to me like a means of sustaining its social licence [sic] to extract the gas and oil that will destroy our lives.").

188. *See* Harvey M. Silets & Susan W. Brenner, *The Demise of Rehabilitation: Sentencing Reform and the Sanctioning of Organizational Criminality*, 13 AM. J. CRIM. L. 329, 367 (1986).

189. *See* Brent Fisse, *Reconstructing Corporate Criminal Law: Deterrence, Retribution, Fault, and Sanctions*, 56 S. CAL. L. REV. 1141, 1204–06 (1983).

190. *See* Irwin Schwartz, *Toward Improving the Law and Policy of Corporate Criminal Liability and Sanctions*, 51 AM. CRIM. L. REV. 99, 112 (2014).

191. *See generally* William D. Smart, Cindy M. Grimm & Woodrow Hartzog, *An Education Theory of Fault for Autonomous Systems*, 2017 PROC. WEROBOT 24–27, http://people.oregonstate.edu/~smartw/library/papers/2017/werobot2017.pdf [https://perma.cc/8EPL-WART]. For a detailed treatment on how bias can arise in algorithms, see Nizan Geslevich Packin & Yafit Lev-Artez, *Learning Algorithms and Discrimination*, *in* RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE 88 (Woodrow Barfield & Ugo Pagallo eds., 2018) (ebook).

192. *See* Kate Crawford, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES (June 25, 2016), https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html [https://perma.cc/FS5W-R5HL (dark archive)].

193. *See* Geistfeld, *supra* note 17, at 1634–36 (discussing algorithm errors caused by programming bugs).

194. *See* Barocas & Selbst, *supra* note 29, at 680–81; Oscar H. Gandy, Jr., *Engaging in Rational Discrimination: Exploring Reasons for Placing Regulatory Constraints on Decision Support Systems*, 12 ETHICS & INFO. TECH. 29, 31 (2010) (discussing how bad data can bias automated systems).

more extensive pre-roll-out testing,[195] regular post-roll-out quality audits,[196] routine run-time compliance layers,[197] effective monitoring,[198] and continuous software updates to address problems as they arise.[199] Programmers also have tools they can use to prove (to themselves or to others) that an algorithm has been applying its rules consistently.[200] Each of these precautions entail costs that, all things considered, corporations would rather avoid. Through the threat of sanction, criminal liability can make taking precaution cheaper than risking violation.

Since the focus of this Article has been to expand the scope of corporate liability, the most pressing policy concern is whether the proposals go too far. For example, the approach offered here does not require any sort of wrongdoing on the part of the corporation aside from the algorithmic misconduct itself. That may seem like a retributively inappropriate form of vicarious liability[201]: How can a corporation deserve punishment if it has done nothing wrong?[202]

195. *See* Geistfeld, *supra* note 17, at 1623, 1651–54; *see also* DAVE CLIFF & LINDA NORTHOP, GOV'T OFFICE FOR SCI., THE GLOBAL FINANCIAL MARKETS: AN ULTRA-LARGE-SCALE SYSTEMS PERSPECTIVE 19–20 (2012) (discussing the need for testing trading algorithms using simulations).

196. *See* B. Bodo et al., *Tackling Algorithmic Control Crisis—The Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents*, 19 YALE J.L. & TECH. 133, 142–44 (2017) (describing audits of algorithms); James Guszcza et al., *Why We Need To Audit Algorithms*, HARV. BUS. REV. (Nov. 28, 2018), https://hbr.org/2018/11/why-we-need-to-audit-algorithms. *See generally* Shlomit Yanisky-Ravid & Sean K. Hallisey, *Equality and Privacy by Design: A New Model of Artificial Intelligence Data Transparency Via Auditing, Certification, and Safe Harbor Regimes*, 46 FORDHAM URB. L.J. 428, 434–35 (2019) (advocating for a system of auditing and certification of data and artificial intelligence in order to encourage transparency).

197. *See* Louise Dennis et al., *Formal Verification of Ethical Choices in Autonomous Systems*, 77 ROBOTICS & AUTONOMOUS SYSS. 1, 1–2 (2016) (discussing formal verification); Felippe Meneguzzi & Michael Luck, *Norm-Based Behaviuor Modification in BDI Agents*, 2009 PROC. 8TH INT'L JOINT CONF. ON AUTONOMOUS AGENTS & MULTI-AGENT SYSS. 177, 177.

198. King et al., *supra* note 81, at 110–12 (discussing four possible monitoring mechanisms for algorithms).

199. *See generally* NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., FEDERAL AUTOMATED VEHICLES POLICY: ACCELERATING THE NEXT REVOLUTION IN ROADWAY SAFETY 16 (2016), https://www.hsdl.org/?view&did=795644 [https://perma.cc/6BWR-YTTR] (envisioning manufacturers of self-driving cars will update software regularly to improve safety).

200. Kroll et al., *supra* note 117, at 662–72 (describing available mechanisms). Without these mechanisms, verification after the fact can be difficult for a host of technical and legal reasons. *See generally* Amanda Lewindowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 589–90 (2018) (examining whether the fair use doctrine can allow AI creators to avoid inconsistencies caused by copyright law).

201. The Association for Computing Machinery has proposed something analogous for individuals, namely that they be held accountable for the decisions made by the algorithms they use. ASS'N FOR COMPUTING MACH., U.S. PUB. POLICY COUNCIL, STATEMENT ON ALGORITHMIC TRANSPARENCY AND ACCOUNTABILITY 2 (2017), https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf [https://perma.cc/YVT6-5GSK].

202. *See* Lemley & Casey, *supra* note 17, at 1313–15 (considering a similar question).

1. Vicarious Liability for Wayward Algorithms

This challenge is not unique to extended mind theory and algorithmic misconduct. All corporate liability, including criminal liability under current law, seems to be vicarious. Corporations cannot act on their own;[203] they must act through the employees and (now) algorithms that run them. Respondeat superior assigns fault from employees to corporations, and it does so without requiring any additional fault, like negligent hiring practices or ex-post ratification of the misconduct, on the part of the corporation.[204] The doctrines proposed above would be no different. So far as the accusation of retributive unfairness is concerned, the response on behalf of respondeat superior and extended mind theory is the same: the challenge relies on a conceptual mistake. The response assumes that there is a gap between the corporation doing something wrong and its employees or algorithms doing something wrong. However, employees and algorithms are parts of corporations.[205] So employee and algorithmic wrongs *are* corporate wrongs. The liability is not really vicarious after all.[206]

There is also a deterrence-based rationale that animates respondeat superior and extended corporate mind theory. Even though corporations can never guarantee that their employees will behave on the job,[207] corporations are in the best position to mitigate the risks of misbehavior.[208] By threatening to punish a corporation whenever one of its employees does something wrong, respondeat superior incentivizes corporations to implement compliance protocols such as[209] additional screening at hiring, better employee training, more effective employee monitoring, more open reporting channels, and stricter disciplinary responses.[210] The same is true of holding corporations responsible for algorithmic misconduct. By holding corporations responsible for

203. Laufer, *Corporate Bodies*, *supra* note 166, at 652 ("A corporation can only act through an agent . . . .").

204. Some theorists proposed that these sorts of corporate deficiencies could form an alternate framework for establishing corporate fault. *See, e.g.*, Fisse, *supra* note 189, at 1200 (discussing both reactive and proactive fault).

205. *See* N.Y. Cent. & Hudson River R.R. v. United States, 212 U.S. 481, 492–93 (1909) (noting that a corporation is made up of its officers and agents).

206. *See* Am. Med. Assoc. v. United States, 130 F.2d 233, 253 (D.C. Cir. 1942) ("When a corporation is guilty of a crime it is because of a corporate act, a corporate intent . . . . The fact that a corporation can act only by human agents is immaterial."), *aff'd*, 317 U.S. 519 (1943).

207. *See* Cindy R. Alexander & Mark A. Cohen, *Why Do Corporations Become Criminals? Ownership, Hidden Actions, and Crime as an Agency Cost*, 5 J. CORP. FIN. 1, 6 (1999) ("[T]he occurrence of crime realistically depends on a variety of influences beyond management's control."); Schwartz, *supra* note 190, at 112.

208. Jennifer Arlen, *The Failure of the Organizational Sentencing Guidelines*, 66 U. MIAMI L. REV. 321, 332–33 (2012) ("[C]orporations often are the lowest-cost providers of many forms of policing.").

209. *See id.* at 332–34.

210. *See* Tanina Rostain, *General Counsel in the Age of Compliance: Preliminary Findings and New Research Questions*, 21 GEO. J. LEGAL ETHICS 465, 466–67 (2008).

the algorithms they use, the law can incentivize corporations to do a better job designing, monitoring, and correcting their algorithms.

One might worry that the same reasoning will not work if, as is often the case, a corporation hires a more experienced technology firm to design its algorithms.[211] In that case, the most direct way to prevent crime might be to target the incentives of the technology firm. However, holding corporate end users liable can accomplish the same result. Corporations will undoubtedly pass the costs of algorithmic misconduct onto technology firms through indemnification agreements, thereby forcing the technology firms to internalize the risk of misconduct.[212] If they bear the financial risk when their algorithms misbehave, technology firms will take more efficient precautions in designing and testing their products. It is administratively easier to hold corporate end users liable rather than going to the technology firms directly. That removes the courts from the messy business[213] of apportioning liability between technology firms (for design error)[214] and corporate end users (for user error).[215]

2. Vicarious Liability for Others' Information

There is another respect in which extended mind theory could lead to what may seem like an overbroad expansion of corporate liability. Recall that under extended mind theory if a person bears the right functional relationship to some information, she counts as knowing it, regardless of where or how the information is stored. In a digitally connected world, this could end up being *very* inclusive. Consider what that means for natural people. Barry is counted as knowing the information in his diary because he could easily use it to find

211. *See* Lemley & Casey, *supra* note 17, at 1352 ("Robots are composed of many complex components . . . often designed, operated, leased, or owned by different companies.").

212. *See* Pamela H. Bucy, *Corporate Ethos: A Standard for Imposing Corporate Criminal Liability*, 75 MINN. L. REV. 1095, 1146 (1991) ("[M]ost corporations use at least one form of compensation, indemnification, in a way that encourages corporate crime . . . ."); David R. Cohen & Roberta D. Anderson, *Insurance Coverage for "Cyber-Losses,"* 35 TORT & INS. L.J. 891, 926 (2000) ("Directors' and officers' [("D&O")] insurance policies afford coverage for the defense and indemnification costs of directors and officers sued in connection with discharge of their corporate duties. A typical D&O policy insures against any 'loss' arising out of a 'wrongful act.'").

213. *See* Lemley & Casey, *supra* note 17, at 1352–53 ("Robot designers, owners, operators, and users will, of course, fight over who bears true legal responsibility for causing the robot to behave the way it did. And these complex distinctions don't even account for the role of third parties causing robots to behave in adverse ways . . . .").

214. Assuming the mistake on the design side, it may still be difficult to locate where. *See* David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 125 (2014) (discussing the difficulties in determining who is responsible for AI errors).

215. *See* MATILDA CLAUSSÉN-KARLSSON, ARTIFICIAL INTELLIGENCE AND THE EXTERNAL ELEMENT OF THE CRIME 22 (2017), http://oru.diva-portal.org/smash/get/diva2:1115160/FULLTEXT01.pdf [https://perma.cc/9F29-WQNR] (discussing involvement of both users and producers of AI); McAllister, *supra* note 123, at 2550.

the café. But what if, instead of writing the directions in his diary, Barry looked them up on his cellphone after he set off. Assuming Barry still bore the right functional relationship to the information, he would count as knowing it. Since he always had easy access to the information through his phone (he could have looked it up any time), he should count as having known it even before he pulled the phone out of his pocket. The fact that the data was actually stored on a server farm hundreds of miles away would not affect the analysis. What is true of the directional information could equally be true of any other information accessible through Barry's data connection. For example, Barry, cell phone in hand, may also count as knowing everything on Wikipedia, even articles he has yet to read. That can lead to some very weird results, e.g., that what Barry knows is constantly being altered without his awareness as Wikipedia editors add, delete, and change content.

Even for many extended mind enthusiasts, this seems a step too far.[216] To constrain the extended mind thesis, some theorists propose an additional criterion for evaluating whether a subject counts as knowing external information: the subject must have previously endorsed the information.[217] This would effectively restrict the scope of information a subject can know to information that had previously been routed through him (i.e., through his brain). With this new restriction, Barry would still count as knowing the directions in his diary since he read them on his computer and then wrote them himself. But he would no longer count as knowing everything on Wikipedia, because he had never previously been cognizant of most of it.

Some similarly restrictive criterion may be appropriate in the corporate context. In the course of fulfilling their duties, corporate employees often use proprietary databases owned, maintained, and operated by third parties.[218] The same is true of corporate algorithms.[219] Loan approval platforms, for example, automatically draw on databases that credit rating agencies maintain.[220] If information on third-party databases is sufficiently integrated into a corporation's algorithmic decisionmaking, the corporation could qualify as knowing it. And what the corporation knows would be in a constant state of flux as the third party maintaining the database changes its content. Such a result could be worrying from a criminal justice perspective. Since corporations cannot directly control the information on third-party servers, holding

216. *See* Clark & Chalmers, *supra* note 43, at 17.

217. *See id*.

218. *See* Pittman, *supra* note 9, at 768–79; *see also* Hillary Hellmann, *Acknowledging the Threat: Securing United States Pipeline SCADA Systems*, 36 ENERGY L.J. 157, 161 (2015).

219. *See* Pittman, *supra* note 9, at 767–69 (discussing third party resources generally).

220. *See, e.g.*, *Comprehensive Credit Decisioning Software*, ACTICO, https://www.actico.com/solutions/loan-origination-decisioning/credit-decision-platform/ [https://perma.cc/Y47T-D2QP] ("The Credit Bureau Gateway provides seamless integration with external data providers, such as credit bureaus.").

corporations to account for it could seem unfair. Risk averse corporations might refrain from giving their algorithms access to third-party information, even when access would otherwise be cost-effective.

Assuming it would be desirable to restrict the scope of the extended corporate mind, a variety of limiting principles are available. One very restrictive condition would require that information housed in a data system, or facts from which the information is inferable, must have been placed there by an employee who previously endorsed that information. This directly parallels the restriction proposed by extended mind theorists for natural people,[221] effectively amounting to a requirement that the corporation would qualify under respondeat superior as once having known the information. A range of weaker conditions could turn on the degree of control corporate employees have over the information, requiring anything from aggressive continuous monitoring to occasional quality control sampling. A further question would be whether the corporation must actually exercise that control or whether it is enough that the corporation merely had the power (legal, practical, or otherwise) to do so.

There may, however, be good reason to forego any additional restricting conditions on corporate knowledge. The intuitions that drive the search for a limit in the individual context are weaker when it comes to corporations. There is an intuitive understanding of what it means for a human subject to "previously endorse" information. Human subjects are spatiotemporally constrained biological units. The question posed by the extended mind thesis is whether their cognition might not extend beyond their spatiotemporal constraints. Corporations, however, are by their nature spatiotemporally distributed subjects. There is no equivalent of the corporate cranium to point to. Since there is no strong intuition against extending the corporate mind to remote data systems managed by other entities, the retributive case for a limiting condition is weak.

There may also be a strong deterrence-based argument for rejecting a limiting condition. One animating worry behind this Article is that, under current law, corporations can insulate themselves from liability by offloading operations from employees to algorithms. But a version of this worry might rearise if corporations can exploit a limiting condition by offloading operations from some algorithms and databases to others. Suppose, for example, a limiting condition required some kind of employee monitoring of information, like random quality control audits. A corporation could manage its liabilities by offloading the information to remote databases maintained by another entity. Furthermore, it may be most efficient in many circumstances to hold corporations to account for all information they routinely access and use. This

221. *See* Clark & Chalmers, *supra* note 43, at 17.

would give corporations incentives to maintain quality controls and to pressure third-party information custodians to do the same.[222] "[T]he safest way to secure care is to throw the risk upon the person who decides what precautions shall be taken."[223] Of course, allocating risk brings transaction costs and other potential barriers to business relations, all of which dampen innovation and economic progress. It is an open empirical question whether something like the fourth requirement would ultimately help or hurt in the corporate context. At least where the potential social stakes are high—as they often are with knowledge-based civil and criminal violations—perhaps the law should require corporations to be exacting about the quality of their information, regardless of where it comes from.

CONCLUSION

Automation is the future for many corporations. That future will make corporations faster and cheaper, but it will not eliminate corporate harm. The law, as it presently stands, will soon find itself without any tools to address broad swathes of corporate misconduct. Most corporate liability requires corporate mental states—like knowledge of falsity or intent to defraud—which the law presently defines in terms of employee mental states. But when algorithms run the corporate show, employee mental states, and hence corporate liability, are out of the picture. This Article proposes a solution that leverages the current framework for corporate liability. Drawing on themes from contemporary philosophy and cognitive science, it shows that minds are not limited by traditionally presumed boundaries. A range of external cognitive aids fulfill brain-like roles for human beings. According to the extended mind thesis, these aids form part of human minds. Similarly, a range of algorithmic aids are coming to fulfill employee-like roles for corporations. Correspondingly, this Article offers discrete legal reforms for recognizing that corporate minds can extend to these too. The basic idea is that corporations that use algorithms to fulfill employee roles should be treated as having the same mental states that corporations using employees to fulfill those roles would have. This reform would prevent opportunistic corporations from limiting their liability risk by offloading operations from employees to algorithms.

It should now be clear that the proposal satisfies this Article's minimalist ambition—very little about current law would need to change. The proposal draws heavily on corporate law's current liability framework. Indeed, the law's fiction of corporate personhood is a crucial motivation for adapting extended

222. The rationale here parallels the case for holding manufacturers of self-driving cars liable for injuries caused by third-party hackers. *See* Geistfeld, *supra* note 17, at 1690.

223. OLIVER WENDELL HOLMES, JR., THE COMMON LAW 117 (Little, Brown & Co. 1923) (1881).

mind theory from natural people to the corporate context. The move only works on the fictionalizing assumption that corporations are people with minds like ours. As importantly, this Article nowhere assumed, as the law does not, that algorithms have minds or can be responsible. Under the extended mind thesis, the hypothesis is not that the external cognitive aids have their own independent mental states.[224] Barry's diary did not know how to get to the café, even if Barry (with the directions written in his diary) did. Analogously, the claim here is that algorithms can form part of the corporate mind, not that they have minds of their own. Corporations can be directly liable for the things they decide and do, even when they use AI to make those decisions and take those actions.

224. Andy Clark, *Coupling, Constitution, and the Cognitive Kind*, *in* THE EXTENDED MIND, *supra* note 109, at 81, 83.