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

Mihailis E. Diamantis
The University of Iowa
College of Law
mihailis-diamantis@uiowa.edu

Working Paper

ComplianceNet is an international network of scholars from across the social and behavioral sciences who study compliance, broadly defined as the interaction between rules and individual and organizational behavior. The network publishes a working paper series that offers the latest work in this field. The papers are unpublished drafts or pre-published versions of publications. Submissions can be sent to the editors, Melissa Rorie, melissa.rorie@unlv.edu, Benjamin van Rooij, bvanrooij@law.uci.edu, and Yuval Feldman, Yuval.Feldman@biu.ac.il

ComplianceNet.org
Acknowledgements: For invaluable feedback at various stages, I owe special thanks to 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), and the Business Law and Technology Roundtable (Michigan State University College of Law).
The Extended Corporate Mind: When Corporations Use AI to Break the Law
Mihailis E. Diamantis

Abstract

Algorithms may soon replace employees as the leading cause of corporate harm. 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 decision-making. 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 the thesis to corporate law. It motivates and proposes a doctrinal framework for extending the corporate mind to the algorithms that are increasingly integral to corporate thought. The law needs such an innovation if it is to hold future corporations to account for their most serious harms.

Keywords: corporate crime; artificial intelligence; algorithmic misconduct; corporate compliance; mens rea, jurisprudence
TABLE OF CONTENTS

I. The Problem of Algorithmic Corporate Misconduct ......................................................... 4
II. A Minimally Invasive Method ....................................................................................... 8
III. HealthCo and FormBot: A Clarifying Example ........................................................... 12
IV. The Extended Mind Thesis .......................................................................................... 15
V. Extending the Corporate Mind ....................................................................................... 17
   A. Doctrinal Proposal ...................................................................................................... 19
      1. The Analogical Approach ..................................................................................... 19
      2. Using Generalized Criteria .................................................................................. 20
   B. Policy-Based Objections and Restrictions ................................................................. 22
VI. Conclusion ..................................................................................................................... 27

I. The Problem of Algorithmic Corporate Misconduct

Marvin\(^1\) makes investments on behalf of a 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 non-public
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 is good grounds
for a closer look. Since Marvin learned material non-public information in the course of his employment, the law
dictates that SciBank learned it as well.\(^3\) SciBank’s liability would then turn on what else Marvin (and hence
SciBank) knew about the provenance of that information.\(^4\)

But suppose that Marvin is not a human being at all; suppose “Marvin” is one of many algorithmic trading
programs in use today.\(^5\) Here the insider trading inquiry quickly 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.\(^6\)

---

\(^1\) I am grateful to Aaron Ancell, at the Edmond J. Safra Center for Ethics, Harvard University, for this opening conceit.
\(^2\) Bernard Marr, The Revolutionary Way of Using Artificial Intelligence in Hedge Funds – The Case of Aidyia, FORBES (Feb.
hedge-funds-the-case-of-aidyia/#17eb640157ca.
\(^3\) See N.Y. Cen. & Hudson River R.R. Co. v. United States, 212 U.S. 481, 485 (1909); Philadelphia, Wilmington, and
\(^5\) See Marr, supra note 2.
\(^6\) 17 C.F.R. § 240.10b5-1.
The SciBank hypo is not futuristic or idiosyncratic.\textsuperscript{7} Advanced algorithms utilizing big data and artificial intelligence are rapidly reshaping every corner of modern business.\textsuperscript{8} Experts predict that corporate reliance on digital automation will increase exponentially over the coming years.\textsuperscript{9} Algorithms are taking over throughout the corporate hierarchy, from the lowest-level operations—like the systems running Amazon’s box-packing bots\textsuperscript{10}—to the highest—like “Vital,” the algorithm appointed to the board of Deep Knowledge Ventures.\textsuperscript{11}

Algorithms promise to make corporations more efficient\textsuperscript{12} and (perhaps)\textsuperscript{13} more objective,\textsuperscript{14} but they do not remove (or even always reduce)\textsuperscript{15} the possibility that things will sometimes go awry.\textsuperscript{16} Indeed, the speed and


\textsuperscript{13}Mark A. Lemley & Bryan Casey, Remedies for Robots, 82 U. CHI. LAW REV. (forthcoming 2019) (“As robotics and artificial intelligence systems increasingly integrate into our society, they will do bad things.”); Matthew Adam Bruckner, The Promise and Perils of Algorithmic Lenders’ Use of Big Data, 93 CHI.-KENT L. REV. 3, 6 (2018); Mark A. Geistfeld, A
geographic reach of algorithmic processes means that when things go wrong, they can go really wrong for a lot of people in a lot of places at once.\textsuperscript{17} Real life examples of corporate algorithmic harm that merit a searching liability inquiry include:\textsuperscript{18}

- A lender’s automated platform approves mortgages in a fashion that has a discriminatory racial impact but might also have a business justification.\textsuperscript{19}
- Competing retailers’ pricing algorithms set prices at matching, super-competitive levels.\textsuperscript{20}
- A delivery company’s self-driving truck strikes a jay-walking pedestrian.\textsuperscript{21}

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.

The move toward automation does not alter the fact that discrimination, price fixing, and reckless driving leave victims in their wake.\textsuperscript{22} These victims, 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. Once corporations remove employees from the picture, corporate liability is the \textit{only} means of redress. Employees are accountable for their own misconduct, whether on or off the job.\textsuperscript{23} Algorithms, however, are not subject to suit.\textsuperscript{24}

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. Under current law, corporate liability in each of the above cases would require (and I assume below should require)\textsuperscript{25} evidence of a culpable corporate mental state:

\begin{itemize}
  \item A lender’s automated platform approves mortgages in a fashion that has a discriminatory racial impact but might also have a business justification.
  \item Competing retailers’ pricing algorithms set prices at matching, super-competitive levels.
  \item A delivery company’s self-driving truck strikes a jay-walking pedestrian.
\end{itemize}

---


\textsuperscript{18} A growing scholarly literature discusses others. See, e.g., Katyal, supra note 12.


\textsuperscript{22} For a detailed discussion of the harms of algorithmic discrimination, see O’NEIL, supra note 15.


\textsuperscript{25} See infra Part II.
purpose (to discriminate), knowledge (of competitors’ prices), or recklessness (in operating a vehicle). The legal doctrine for attributing mental states to corporations—respondeat superior—defines corporate mental states in terms of employee mental states. Variants of respondeat superior—like the collective knowledge doctrine, which allows courts to aggregate employee knowledge, and the control group test, which restricts respondeat superior to higher ranking corporate employees—only 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. But when corporations using algorithms misbehave in ways that, from the outside, look just as purposeful, knowing, or reckless, current liability doctrines do not apply.

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. Its processors analyze that information, on some estimates, 10,000,000 times faster than any human could. 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. The legal loophole left by respondeat superior incentivizes an unpalatable form of corporate gamesmanship. Corporations keen to manage their liabilities 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

28 See, e.g., IOWA CODE ANN. § 321.777 (West) (defining offense of “reckless driving”).
31 See MODEL PENAL CODE § 2.07(1)(c).
32 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) (critiquing respondeat superior).
The Extended Corporate Mind

their embrace of automation. This exacerbates the risk that corporations will turn to algorithms prematurely, before the technology has been sufficiently tested for socially responsible use.38

What the law needs is a doctrinal framework for extending its conception of the corporate mind beyond the employees whose shoes algorithms are coming to fill. Psychologists and philosophers have recently addressed 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.39 The so-called “extended mind thesis” states that the human mind reaches beyond the brain to encompass external cognitive aids—like diaries or cellphones—that help the brain do its work.40 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.41 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 Article begins by clarifying important aspects of its methodology (Part II) and offering a focusing hypothetical scenario (Part III). In its chief substantive contribution, the Article introduces the extended mind thesis (Part IV) and shows in detail how to adapt it as a doctrine for addressing corporate algorithmic misconduct (Part V). The Article concludes by reflecting on broader implications of the extended corporate mind (Part VI).

II. 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. The basic idea is to demolish and rebuild. This Article’s more modest ambition is to use a scalpel. It aims to solve the problem of algorithmic corporate misconduct by making the smallest possible doctrinal incision. What it may sacrifice in terms of grandiose vision, surgical intervention makes up in terms of feasibility. It does this by leveraging existing frameworks and doctrines of corporate liability to address cases 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.42 This feature is the present proposal’s distinctive advantage over others.

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

38 Microsoft President and Chief Legal Officer Brad Smith has remarked, “We don’t want to see a commercial race to the bottom. Law is needed.” Metz, supra note 15. See Frank Pasquale, Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society, 78 OHIO ST. L.J. 1243, 1244 (2017).
40 Id. at 12-14 (discussing an example involving a diary).
41 Id. at 13.
42 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 (1995); GUIDO CALABRESI, A COMMON LAW FOR THE AGE OF STATUTES 4 (1982); 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.”); Faith, Hope, and Rationality: Or, Public Choice and the Perils of Occam's Razor, 28 Fla. St. U. L. Rev. 109, 110-11 (2000) (“[P]ublic choice is ‘appealing in its parsimoniousness,’ . . . Even if we had the ability to dismantle the entire national regulatory apparatus, we have neither the will nor the desire to do so.”).
Some scholars believe that corporate culpability is such a nonsensical notion\textsuperscript{43} that the law would be better without it.\textsuperscript{44} They may be right. After all, corporations are just fictional agents.\textsuperscript{45} Perhaps the best solution to algorithmic corporate misconduct would begin by prodding 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. 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\textsuperscript{46} and often makes good policy sense.\textsuperscript{47} More to the point, scrapping corporate culpability is a practical nonstarter. The public broadly supports holding corporations civilly and criminally accountable.\textsuperscript{48} Consequently, the law of corporate culpability is socially entrenched and politically bulletproof. Furthermore, the legal edifice built around corporate culpability is centuries old\textsuperscript{49} and growing.\textsuperscript{50} Taking it apart would be a massive jurisprudential undertaking.

No feature of the law’s framework for corporate culpability is more integral than the fiction of corporate personhood.\textsuperscript{51} According to this social construct, corporations fit into the law’s liability mechanisms just as other “people” do.\textsuperscript{52} Any statute that defines a civil or criminal violation simultaneously sets out elements of liability for both individuals and corporations. Whatever act a statute requires, it requires of real and fictional people alike. Where, as is most often the case,\textsuperscript{53} liability also turns on the person having a concurrent mental state, both individual and corporate defendants must satisfy it for them to be liable. Some theorists think that because corporations are not really people, the law should abandon its pretense to the contrary.\textsuperscript{54} These theorists argue that the law’s present reliance on corporate mental states makes for ineffective justice\textsuperscript{55} and suboptimal prevention.\textsuperscript{56} Removing any reference to corporate mental states would convert all corporate liability into strict liability.

\begin{thebibliography}{99}
\bibitem{44} See, e.g., John Hasnas, \textit{The Centenary of a Mistake: One Hundred Years of Corporate Criminal Liability}, 46 AM. CRIM. L. REV. 1329 (2009).
\bibitem{45} Sierra Club v. Morton, 405 U.S. 727, 742–43 (1972) (“The ordinary corporation is a ‘person’ for purposes of the adjudicatory processes.”).
\bibitem{47} Diamantis, supra note 32.
\bibitem{50} Sara Sun Beale, \textit{A Response to the Critics of Corporate Criminal Liability}, 46 AM. CRIM. L. REV. 1481, 1482 (2009).
\bibitem{51} A corporation is an entity “having authority under law to act as a single person.” \textsc{Black’s Law Dictionary}, corporation (9th ed. 2009).
\bibitem{52} See, e.g., The Dictionary Act, 1 U.S.C. §1 (1976) (defining “person” to include “corporation”).
\bibitem{53} 22 C.J.S. \textit{Criminal Law} § 39 (“Strict liability is the exception in our criminal law system, not the rule, and ha[s] a generally disfavored status.”); Restatement (Third) of Torts: Prod. Liab. (1998), Restatement (Third) of Torts: Phys. & Emot. Harm 4 Scope Note (2010) (strict liability is generally limited to torts involving abnormally dangerous activity, possession of animals, and products liability).
\bibitem{55} Albert W. Alschuler, \textit{Two Ways To Think About the Punishment of Corporations}, 46 AM. CRIM. L. REV. 1359, 1392 (2009).
\end{thebibliography}
This Article is not about the advisability of the law’s fiction of corporate personhood. The law can indulge this fiction regardless of whether or not 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. As I have argued elsewhere, the 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. Additionally, despite all the sophisticated economic theory endorsing strict corporate liability, that approach often results in inefficient overdeterrence and over-investment in compliance. 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.

Lastly, this Article will not follow the lead of scholars in law, computer science, and business ethics who propose addressing algorithmic misconduct by holding the algorithms themselves liable. That project is deeply controversial. It is far from clear that algorithms presently do or ever could satisfy the conditions of personhood and culpability. Even theorists who propose a fictionalizing approach to algorithmic personhood (analogous to the law’s fiction of corporate personhood) face two formidable obstacles. First, no one has

---

58 1 Max Weber, Economy and Society (1914) (arguing that corporations are not agents).
59 See, e.g., International Shoe Company v. Office of Unemployment Compensation & Placement, 326 U.S. 310 (1945) (“[T]he corporate personality is a fiction, although a fiction intended to be acted upon as though it were a fact . . . .”).
60 Diamantis, supra note 32, at 2063-64.
61 See, e.g., Richard A. Posner, An Economic Theory of the Criminal Law, 85 Colum. L. Rev. 1193, 1222 (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.”); Daniel R. Fischel & Alan O. Sykes, Corporate Crime, 25 J. Legal Stud. 319, 328 (1996) (“[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.”).
64 By situating itself within this fiction, this Article need not make any controversial assumptions about the true metaphysics of corporations.
69 See generally David Gray, The Nature and Sources of the Law 27-52 (1921) (discussing legal personhood)
70 See Lawrence B. Solum, Legal Personhood for Artificial Intelligences, 70 N.C. L. Rev. 1231 (1992) (comparing the cases for AI personhood and corporate personhood).
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.\textsuperscript{71} 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.\textsuperscript{72} Tay was supposed to teach itself to talk by learning from data it scraped from Twitter.\textsuperscript{73} Within twenty-four hours, internet users had baited Tay with enough corrupting tweets that the bot’s messages became chauvinistic, racist, and anti-Semitic.\textsuperscript{74} It is far from clear, though, that Tay was to blame for the things it said, and not Microsoft or a corrosive Twitterverse.\textsuperscript{75} 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\textsuperscript{76} (short, perhaps, of killing it, as Microsoft did to Tay).\textsuperscript{77} We can jail or fine other “people,” but algorithms lack bodies and pocketbooks.\textsuperscript{78}

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.\textsuperscript{79} Reversing course would require the sweep 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 Sentencing Guidelines provisions on organizations, the law also already has long-standing mechanisms for sanction corporations.\textsuperscript{80} 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.\textsuperscript{81}

Despite the significant difficulties posed by corporate algorithmic misconduct, this Part has hopefully shown that the ambitions of this Article 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. Lastly, it should not require a new body of law establishing the legal personhood or accountability of algorithms.

\textsuperscript{71} Anupam Chandler, The Racist Algorithm, 115 Mich. L. Rev. 1023, 1037 (2017); David A Patterson & John L. Hennessy, Computer Organization and Design 13-16 (5th ed. 2014). 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 (1996).


\textsuperscript{73} Id.

\textsuperscript{74} Damon Beres, Microsoft Chat Bot Goes on Racist, Genocidal Twitter Rampage, HuffPost (Mar. 24, 2016), https://www.huffpost.com/entry/microsoft-tay-racist-tweets_n_56f3e678e4b04c4c37615502.


\textsuperscript{78} See Ryan Abbott and Alex Sarch, Punishing Artificial Intelligence: Legal Fiction or Science Fiction, 52 U.C. Davis L. Rev. (forthcoming 2019) (discussing and ultimately rejecting the possibility of punishing algorithms).

\textsuperscript{79} Thomas Beardsworth & Nishant Kumar, Who to Sue When a Robot Loses Your Fortune, BLOOMBERG (May 5, 2019), https://www.bloomberg.com/news/articles/2019-05-06/who-to-sue-when-a-robot-loses-your-fortune (“Robots are getting more humanoid every day, but they still can’t be sued.”).

\textsuperscript{80} U.S. Sentencing Guidelines ch. 8.

\textsuperscript{81} 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.
III. HealthCo and FormBot: A Clarifying Example

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

HealthCo 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 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 or civil 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 the challenge posed by the problem of algorithmic misconduct. There is no doubt under current law that corporations can violate the False Claims Act. 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. However, under the hypo as formulated, current law dictates no civil or 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 at issue in the hypo. 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, particularly in the corporate context. But criminal law concerns itself with more than efficiency. Criminal law would not fulfill its distinctive condemnatory and powerful deterrent functions if it relied only on strict liability. That approach necessarily treats the innocent dupe, the hapless fool, and the

---

85 See, e.g., United States v. Sain, 141 F.3d 463 (3d Cir. 1998).
87 See Fischel & Sykes, supra note 56, at 326-28.
90 Samuel W. Buell, The Blaming Function of Entity Criminal Liability, 81 IND. L.J. 473, 500-12 (2006)).
calculating villain all alike.91 These differences matter to victims demanding justice and to defendants facing punishment. A system of criminal law that made no reference to mental states would be unrecognizable.92

Third, the mental state at issue for HealthCo is knowledge.93 Every type of mental state has its own distinctive properties and ideally would receive separate treatment. However, there are over one-hundred different types of mens rea among the provisions of the federal criminal code alone.94 There is not space here to address them all. Knowledge is an element of many of the most common corporate crimes.95 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.96

Fourth and last, the hypo 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 first specify the algorithm’s goal and then train the algorithm with a large set of test cases.97 By telling the algorithm in each test case whether or not it attained its goal, the algorithm can eventually learn to succeed on its own.98 For example, engineers might want to design a drone-flying algorithm to take the most efficient route to a target.99 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.100 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.101 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.102 The resulting algorithms become so complicated that

92 This is why Rebecca Crootof, proposing strict state liability for harms caused by warfare AI, turns from criminal law to tort law. War Torts: Accountability for Autonomous Weapons, 164 U. PA. L. REV. 1347, 1387-88 (2016).
93 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). Philosophers have additional requirements for knowledge, the most 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 (1987). I am using “knowledge” in the legal sense consistent with its meaning in the False Claims Act.
95 See Diamantis, supra note 63 (listing examples).
96 Notable exceptions include Mark Rowlands, Consciousness, Broadly Construed, in THE EXTENDED MIND 271 (Richard Menary, ed., 2010), and Mattia Gallotti & Bryce Huebner, Collective Intentionality and Socially Extended Minds, 30 PHIL. PSYCH. 251 (2017).
99 See Lemley & Casey, supra note 16 (discussing the example).
100 See Lemley & Casey, supra note 16 (“[T]he unpredictability inherent to machine learning is also one of its greatest strengths.”).
101 PEDRO DOMINGO, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD (2015); Ryan Abbott and Alex Sarch call this the “Hard AI Crime.” Supra note 78.
102 Id.
programmers analyzing the code afterwards often cannot understand how it works.\textsuperscript{103} Additionally, many algorithms have built-in randomness as an essential part of their design.\textsuperscript{104} Consequently, machine learning algorithms can behave in unintended and unanticipated (and unanticipatable) ways.\textsuperscript{105} 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 unanticipated interaction between these two factors and the real world.\textsuperscript{106} But because of the code’s complexity, problems can arise even if every human involved is fully innocent.\textsuperscript{107}

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.\textsuperscript{108} 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.\textsuperscript{109} 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 hypo is a pure case of algorithmic misconduct.\textsuperscript{110} Otherwise, the worry might arise that violation would really be a case of 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.\textsuperscript{111} In the hypo, if a HealthCo engineer purposely or knowingly designed FormBot to submit fake forms, respondenat superior would attribute the engineer’s purpose or knowledge to HealthCo, thereby satisfying the False Claims Act’s mental state requirement.\textsuperscript{112} It would not matter that the engineer did not herself physically submit the forms.\textsuperscript{113} If the engineer recklessly designed FormBot, willful ignorance might provide a basis for saying she, and hence HealthCo, had constructive knowledge of the fake forms.\textsuperscript{114} What current doctrine cannot handle, and what this Article sets itself to address, are the most worrying


\textsuperscript{105} Lemley & Casey, \textit{supra} note 16 (“[M]uch of the [algorithmic] misconduct that tomorrow’s designers, policymakers, and watchdogs must guard against might not be intentional at all.”).

\textsuperscript{106} Kroll, et al., \textit{supra} note 104, at 693-94.


\textsuperscript{108} Lemley & Casey, \textit{supra} note 16.

\textsuperscript{109} Id.

\textsuperscript{110} It is thus unlike some examples that other criminal scholars have focused on. \textit{See}, e.g., Amanda McAllister, \textit{Stranger than Science Fiction: The Rise of AI Interrogation in the Dawn of Autonomous Robots and the Need for an Additional Protocol to the UN Convention Against Torture}, 101 MINN. L. REV. 2527, 2547 (2017).

\textsuperscript{111} See Hallevy, \textit{supra} note 76 (2010) (discussing different models of criminal liability for crimes involving robots premised on programmer fault).

\textsuperscript{112} See \textsc{Model Penal Code} \textsection 2.02(5).

\textsuperscript{113} \textit{Causing a claim to be submitted falsely explicitly satisfies the civil version of the False Claims Act. 31 U.S.C. \textsection 3729. For criminal law, other liability doctrines fill the gap. See 18 U.S.C. \textsection 2 (“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.”); \textsc{Model Penal Code} \textsection 5.04(1)(a) (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. \textit{See}, e.g., United States v. Bank of New England, 821 F.2d 844, 856 (1st Cir. 1987).

\textsuperscript{114} \textit{See Glob.-Tech Appliances, Inc. v. SEB S.A., 563 U.S. 754, 769 (2011).}
cases where employee misconduct is removed from the picture. A solution for cases of pure algorithmic 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.

IV. The Extended Mind Thesis

It is time to modernize the law’s conception of corporate mentality. The law can only properly hold corporations to account 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 motivates and describes this understanding as applied to natural people. The next Part broadens the theory and shows how to adapt it, 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. Extended mind theorists typically endorse a “functionalist” account of mental states. 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). For example, 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 behavioral outputs (e.g. walking to the freezer). 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. 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. Human neurons clearly can do the job. But so could systems made of complex arrangements of different material, whether organic (think of animal or alien brains) or inorganic (think of 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 put the point as a “Parity Premise:"

[If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting it as part of the cognitive process, then that part of the world is . . . part of the cognitive process.]

---

116 See Barocas & Selbst, supra note 26, 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, 856, 884-85 (2017).
117 Clark & Chalmers, supra note 39, at 14.
120 BRIAN LOAR, MIND AND MEANING (1981).
121 Hilary Putnam, Psychological Predicates, in ART, MIND, AND RELIGION 37 (W.H. Capitan and D.D. Merrill eds., 1967); Menary, supra note 118, at 6 (“[I]t is not the physical properties that matter to [extended mind theory], however, but the functionality of the process.”).
122 Clark & Chalmers, supra note 39, at 8.
Most typically, the systems underwriting human mental states are located within the skull. But a person whose brain protruded beyond her skull could still have mental states in the protrusion. This could be true even if the protrusion was quite distal, connected by long neurons. All that matters is that the protrusion has everything it needs (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:

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 Alzheimers 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.

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. Because Alice and Barry share the same functional relationship between inputs and outputs, extended mind theorists conclude that since the former knew how to get to the café, so did latter.

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, 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. 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 diary pages on which they 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, 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
3. The subject can easily access the information.


Extended mind theorists call this “causal coupling” and set out criteria for it. Menary, supra note 118, at 3.

Andy Clark, Mementos Revenge: The Extended Mind Revisited, 16 MIND & LANGUAGE 6-7 (2004).
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 complex systems that seem to be extended cognitive systems.\textsuperscript{126} 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.\textsuperscript{127} Other extended cognitive systems that are made up of multiple individuals, e.g. navigational teams,\textsuperscript{128} large-scale scientific research,\textsuperscript{129} and transactive memory systems.\textsuperscript{130} Cognitive scientists draw on themes reflected in the extended mind thesis to explain phenomena in situated cognition,\textsuperscript{131} robotics,\textsuperscript{132} and child development.\textsuperscript{133} The theory of embodied cognition provides the framework for this perspective: “Many features of cognitive science . . . are deeply dependent on characteristics of the physical body . . . [which] play[] a significant causal role . . . in the agent’s cognitive processing.”\textsuperscript{134} The basic idea is that minds are not so much tools for “thinking,” but tools for doing things in the world.\textsuperscript{135} From there, it is a short step to the extended mind thesis.\textsuperscript{136}

V. Extending the Corporate Mind

If the extended mind thesis is true, the traditional boundaries on 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,\textsuperscript{137} parallel arguments should work for corporate people too. The starting premise is that corporations have minds. The law takes care

\textsuperscript{126} I should also note that this view is far from uncontroversial in cognitive science. \textit{See, e.g.}, Robert D. Rupert, \textit{Cognitive Systems and the Extended Mind} (2009); Fred Adams & Kenneth Aizawa, \textit{Why the Mind is Still in the Head, in The Cambridge Handbook of Situated Cognition} 78 (Philip Robbins & Murat Aydede, eds., 2009).


\textsuperscript{130} Transactive memory is a system in which a group of people encodes, stores, and retrieves data. \textit{See} Daniel M. Wegner, \textit{A Computer Network Model of Human Transactive Memory}, 13 SOC. COGNITION 319 (1995).

\textsuperscript{131} Lucy A. Schuman, \textit{Plans and Situated Actions} (1987).

\textsuperscript{132} Randall D. Beer, \textit{Intelligence as Adaptive Behavior} (1989).

\textsuperscript{133} Esther Theelen & Linda B. Smith, \textit{A Dynamic Systems Approach to the Development of Cognition and Action} (1994).


\textsuperscript{136} See Rupert, supra note 126.

\textsuperscript{137} One fascinating article argues that group minds could be formed from the extension of individual minds to other individual minds. Deborah Perron Tollefsen, \textit{From Extended Mind to Collective Mind}, 7 COG. SYS. RESEARCH 140, 141 (2006). On Tollefsen’s view the group mind is the result of the extension, not (as I propose here) the mind which is extended. She explicitly excludes AI from her view. \textit{Id.} at 146.
of that premise by directing its participants to assume corporations are people. Corporate mental states, the law presently tells us, reside within the heads of employees. Given that initial reference point, 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. The Parity Premise, quoted in the previous section, could easily adapt to the corporate context:

If a part of the world functions as a corporate process which, were it performed by an employee, the law would accept it as part of the corporate mind, then that part of the world is part of the corporate mind.

Significantly, that “part of the world” could be smart algorithms 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 way what would be called intelligent if a human were so behaving.”

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. It is solely by dint of legal fiat that corporations have minds at all or that they share the mental states of their employees. 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 aids. That alternative is grounded in commonsense biology—the mind is limited by the brain. But a plausible limiting principle is much harder come by where the corporate mind is concerned. Corporations have no brains of their own.

The remainder of this Part further motivates the extended mind thesis for corporations and discusses how to integrate it into the law of corporate liability. Following two accounts of what the positive doctrine might look like, the Part takes up policy considerations that could call for pragmatic limits on how far the corporate mind extends.

---

138 United States v. A&P Trucking Co., 358 U.S. 121, 206 (1958) (“[I]t is elementary that such impersonal entities can be guilty of ‘knowing’ or ‘willful’ violations of regulatory statutes through the doctrine of respondent [sic] superior.”).

139 The NHTSA seems open to a similar sort of functional reasoning. See Letter from Paul A. Hemmingsbaugh, Chief Counsel, Nat’l Highway Traffic Safety Admin., to Chris Urmson, Director, Google, Inc. Self-Driving Car Project (Feb. 4, 2017), (“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.”).

140 See also Clark, supra note 125, at 9.


142 Trustees of Dartmouth Coll. v. Woodward, 17 U.S. 518, 636 (1819) (“A corporation is an artificial being, invisible, intangible, and existing only in contemplation of law.”).

143 And perhaps some psychological projection on our part. See Diamantis, supra note 32, at 2077-80.


145 A sophisticated version of this challenge is the so-called “coupling-constitution fallacy.” See FREDERICK ADAMS & KENNETH AIZAWA, THE BOUNDS OF COGNITION chs. 5-6 (2010).

146 For similar reasons, the objection to the extended mind thesis that people lack direct access to information in extended systems, see John Preston, The Extended Mind, The Concept of Belief, and Epistemic Credit, in THE EXTENDED MIND 355 (Richard Menary, ed., 2010), do not apply to corporations.

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. Stated abstractly, the approach shows that a person, P₁, knows some external information, I₁, by comparing him to a second person, P₂, who has a relevantly similar functional relationship to similar information, I₂. If it is clear that P₂ knows I₂, then P₁ must know I₁. That is the upshot of defining mental states in terms of their functional role—same function, same mental state.

The analogical approach works slightly differently for corporate people. Where P₁ is a natural person (like Barry), the readiest P₂ with clear knowledge is someone who (like Alice) had the information stored in her brain. However, where P₁ is a corporate person with digitally-stored I₁, the analogical approach calls for a different sort of P₂ comparator. Under respondeat superior, the clear-cut case of corporate knowledge is one where the information is stored in an employee brain. So P₂ should be a corporation that engages in similar behavior, but where it is clear that an employee knows I₂. The success of the analogy turns on two factors: 1) whether the functional relationships between P₁ and I₁ and P₂ and I₂ are relevantly similar, and 2) how obvious it is that P₂ (i.e., an employee at P₂) knew I₂.

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

It bears emphasizing that plaintiffs and prosecutors using the analogical approach would still need to satisfy their relevant burdens of proof: preponderance of the evidence and proof beyond a reasonable doubt, respectively. It would not be enough for them simply to stipulate in their comparison case, P₂, that a corporate employee knew I₂. That would just demonstrate that there is some analogous P₂ who knows I₂, even if P₁ and P₂ are otherwise relevantly similar. The strongest implication factfinders could draw from the case is the mere

---

148 See supra pp. [[]].
149 Some scholars discussing negligence have proposed different standards for employees and mental states, e.g. that the standards for negligence in AI should be twice as permissive as those for human individuals. See Geistfeld, supra note 16, at 1679.
150 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 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.
152 See, e.g., William S. Laufer, Corporate Bodies and Guilty Minds, 43 Emory L.J. 647, 701 (1994).
possibility that the corporate defendant, P₁, knew I₁. Strategically, plaintiffs and prosecutors should instead present a P₂ that had access to the relevant information, behaved similarly to P₁, and did so only using employees. Plaintiffs and prosecutors should not dictate that one of P₂’s employees actually knew the information. To satisfy the burden of proof for liability, the factfinders would have to infer that P₂ (i.e. an employee within P₂) most likely knew (civil law) or must have known (criminal law) the relevant information. That would show not just that there is some analogous P₂ that knew the information, but that any analogous P₂ probably did. The implication that would follow is that P₁ 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 loquitor arguments—i.e. that the sort of accident at issue does not ordinarily occur without negligence—have an identical logical structure. The plaintiff effectively argues that in any relevantly similar case, the person who caused the accident most likely behaved negligently. 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 prosecutor cannot show who. 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 hypo. 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 clear it is that some employee in the comparison case knew the information. Like so much in the factfinding process, the analysis in most cases will not be clear cut. More facts would be needed in the HealthCo hypo before a resolution started to crystallize. It may turn out, for example, that the misstatements on the 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 corporations bear to information that current law clearly treats them as knowing. Respondeat superior defines

---

156 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: PRODUCTS LIABILITY § 3 cmt. B.
157 See Developments in the Law--Corporate Crime: Regulating Corporate Behavior Through Criminal Sanctions, 92 HARV. L. REV. 1227, 1248 (1979); see, e.g., The President Coolidge (Dollar S.S. Co. v. United States), 101 F.2d 638, 640 (9th Cir. 1939).
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.\textsuperscript{158} 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 and that they use to direct corporate operations would satisfy the criteria.

Respondeat superior has two additional requirements for attributing employee knowledge to corporations: that the employee knows the information within the scope of her employment and uses it with an intent to benefit her employer. These two requirements, however, have been so weakened by the courts that it is questionable whether the generalized criteria developed here 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.\textsuperscript{159} And an employee counts as intending to benefit her employer even when that intent is subsidiary,\textsuperscript{160} hypothetical,\textsuperscript{161} or ineffective.\textsuperscript{162}

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.\textsuperscript{163} 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 and 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.

\textsuperscript{158} See supra pp. [__].
\textsuperscript{159} United States v. Hilton Hotels Corp., 467 F.2d 1000, 1004 (9th Cir. 1972).
\textsuperscript{160} United States v. Automated Med. Labs., Inc., 770 F.2d 399, 407 (4th Cir. 1985).
\textsuperscript{162} Old Monastery Co. v. United States, 147 F.2d 905, 908 (4th Cir. 1945).
\textsuperscript{163} Sun-Diamond, 138 F.3d at 970.
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 criteria (automatically endorses the information). 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 receives 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 less likely to count as knowing the information.

It is worth pausing briefly to note that the four criteria initially seem to generate attractive 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. They each provide a theoretical justification and doctrinal scaffold for stretching the law’s current liability regimes to cases 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 the civil context too.

The foremost policy goals in corporate criminal law are familiar to criminal law more broadly, namely retribution and deterrence. Retribution may initially seem an odd fit for corporate criminal law since corporations are not ordinary moral agents. However, retributive sentiments are a strong driver behind corporate criminal law. 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. 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. Illustrative is the clarion call

---

164 Clark, supra note 125, at 9 (describing a similar situation involving a human and his diary).
168 Albert W. Alschuler, Two Ways To Think About the Punishment of Corporations, 46 AM. CRIM. L. REV. 1359, 1392 (2009); Baker, supra note 54, at 350.
169 Baer, supra 48, at 162.
171 See Diamantis, supra note 32, at 2077-80.
for justice against corporations that have pushed addictive opioids onto desperate consumers\textsuperscript{172} or have destroyed delicate environmental habitats.\textsuperscript{173} 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.\textsuperscript{174} The law can do this by using the threat of sanction to induce corporations to take more care in how they run their businesses. Ordinarily, this means corporations will implement additional compliance programs for things like employee training and monitoring.\textsuperscript{175} 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 follow the law either),\textsuperscript{176} software engineers can take steps to reduce the probability that the algorithm will misbehave.\textsuperscript{177} These steps include: diversifying the body of engineers writing algorithms,\textsuperscript{178} more careful initial programming,\textsuperscript{179} more mindful selection of training data sets,\textsuperscript{180} more extensive pre-roll-out testing,\textsuperscript{181} regular post-roll-out quality audits,\textsuperscript{182} routine run-time compliance layers,\textsuperscript{183} effective monitoring,\textsuperscript{184} and continuous software updates to

\textsuperscript{172} See, e.g., Maia Szalavitz, \textit{Big Pharma's Opioid Greed Was Even Worse Than We Thought}, VICE (Sept. 13, 2018) https://www.vice.com/en_us/article/7xj97q/big-pharmas-opioid-greed-was-even-worse-than-we-thought (“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.”).


\textsuperscript{179} See Geistfeld, \textit{supra} note 16, at 1634-36 (discussing algorithm errors caused by programming bugs).


\textsuperscript{181} See Geistfeld, \textit{supra} note 16, at 1623; id. at 1651-54; Dave Cliff & Linda Northrop, \textit{The Global Financial Markets: An Ultra-Large-Scale Systems Perspective}, 2012 MONTEREY WORKSHOP 29 (discussing the need for testing trading algorithms using simulations).


\textsuperscript{184} King, \textit{supra} note 68 (discussing four possible monitoring mechanisms for algorithms).
address problems as they arise.\textsuperscript{185} Programmers also have tools they can use to prove (to themselves or to others) that an algorithm has been applying its rules consistently.\textsuperscript{186} Each of these precautions entail costs that, all things considered, corporations would rather avoid. Through the threat of sanction, the criminal law can make taking precaution cheaper than risking a violation.

Since the focus of this Article has been to expand the scope of corporate liability, the most pressing policy worry 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:\textsuperscript{187} How can a corporation deserve punishment if it has done nothing wrong?\textsuperscript{188}

This challenge is not unique to extended mind theory and algorithmic misconduct. All corporate liability, including criminal liability under current law, is vicarious. Corporations cannot act on their own;\textsuperscript{189} they must act through the employees and (now) algorithms that run them. Respondeat superior assigns fault from employees to corporations, and does so without requiring any additional fault (like negligent hiring practices or ex-post ratification of the misconduct) on the part of the corporation.\textsuperscript{190} 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. It 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.\textsuperscript{191} So employee and algorithmic wrongs are corporate wrongs. The liability is not really vicarious after all\textsuperscript{192}

There is also a deterrence-based rationale that animates respondeat superior and extended mind theory. Even though corporations can never guarantee that their employees will behave on the job,\textsuperscript{193} corporations are in the best position to mitigate the risks of misbehavior.\textsuperscript{194} By threatening to punish a corporation whenever one of its employee does something wrong, respondeat superior gives the corporation incentives to implement compliance protocols such as:\textsuperscript{195} more screening at hiring, better employee training, more effective employee

\textsuperscript{186} Kroll, et al., supra note 104, at 662-672 (describing available mechanisms). Without these mechanisms, verification after the fact can be difficult for a host of technical and legal reasons. See generally Amanda Lewindowksi, How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem, 93 Wash. L. Rev. 579 (2018).
\textsuperscript{187} The Association for Computing Machinery has proposed something analogous for individuals, 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.
\textsuperscript{188} Lemley & Casey, supra note 16.
\textsuperscript{189} Lauf er, supra note 152, at 652 (“A corporation can only act through an agent.”).
\textsuperscript{190} Some theorists proposed that these sorts of corporate deficiencies could form an alternate framework for establishing corporate fault. See, e.g., Brent Fisse, Restructuring Corporate Criminal Law: Deterrence, Retribution, Fault, and Sanctions, 56 S. Cal. L. Rev. 1141, 1200 (1983).
\textsuperscript{192} Am. Med. Assoc. v. United States, 130 F.2d 233, 253 (D.C. Cir. 1942), aff’d, 317 U.S. 519 (1943) (“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.”).
\textsuperscript{195} Id. at 332-34.
monitoring, more open reporting channels, and stricter disciplinary responses. The same is true of holding corporations responsible for algorithmic misconduct. By holding corporations responsible for 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. 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. 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 of apportioning liability between technology firms (for design error) and corporate end users (for user error).

There is another respect in which extended mind theory could lead to what may seem like an over-broad 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 counted as knowing the information in his diary because he could easily use it to find 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 anytime), 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. To constraint the extended mind thesis, some theorists advocate for an additional criterion for when a subject counts as knowing external information: The subject must have previously endorsed the information. That 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 had never previously been cognizant of most of it.

---

197 See Lemley & Casey, supra note 16 (“[R]obots are composed of many complex component . . . often designed, leased, or owned by multiple different companies.”).
199 See Lemley & Casey, supra note 16.
202 Clark & Chalmers, supra note 39, at 17.
Without a similarly restrictive criterion, the radical inclusivity of the extended mind thesis would carry over to the corporate context. In the course of fulfilling their duties, corporate employees often use proprietary databases owned, maintained, and operated by third parties.\(^{203}\) The same is true of corporate algorithms.\(^{204}\) Loan approval platforms, for example, automatically draw on databases that credit rating agencies maintain.\(^{205}\) If information on third-party databases is sufficiently integrated into a corporation’s algorithmic decision-making, 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 corporations to account for it could seem unfair. Furthermore, risk averse corporations could 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 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, 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 it 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. 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 would give corporations incentives to maintain quality controls and to pressure third-party information custodians to do the same.\(^{206}\) “The safest way to secure care is to throw the risk upon the person who decides which precautions shall be taken.”\(^{207}\) 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.

---


\(^{204}\) See Pittman, supra note 8, at 767-68.

\(^{205}\) See, e.g., Actico, Comprehensive Credit Decisioning Software, https://www.actico.com/solutions/loan-origination-decisioning/credit-decision-platform/ (last accessed June 26, 2019) (“The Credit Bureau Gateway provides seamless integration with external data providers, such as credit bureaus.”).

\(^{206}\) 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 16, at 1690.

\(^{207}\) Oliver Wendell Holmes, Jr., The Common Law 117 (1881).
VI. Conclusion

Automation is the future for many corporations. That future will make them 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 extend mind thesis, these aids form part of their minds. Similarly, a range of algorithmic aids are coming to fulfill employee-like roles for corporations. Correspondingly, the 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 remove the opportunity corporations presently have to limit liability risk by offloading operations from employees to algorithms.

It should now be clear that the proposal satisfies the 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 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, the 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.208 Barry’s diary did not know how to get to the cafe, even if Barry (with the directions written in his diary) did. Analogously, the claim here is 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.

---