

Is Corporate Responsibility Relevant to Artificial Intelligence Responsibility?

VIKRAM R. BHARGAVA* AND MANUEL VELASQUEZ**

ABSTRACT

Several theorists have suggested that it is possible to understand and elucidate the moral responsibility of artificial intelligence (AI) agents by using the notion of corporate moral responsibility as a model. We argue that there are differences between corporations and AI agents that make the notion of corporate moral responsibility an inappropriate model for AI agent moral responsibility. We discuss both the moral arguments that purport to show that corporations should or should not be held morally responsible, and the metaphysical arguments that purport to show that a corporation is the kind of entity that can be held morally responsible. We show that these arguments either are wrong or are inapplicable in the context of AI agents.

TABLE OF CONTENTS

INTRODUCTION	830
I. THE MORAL REASONS FOR (AND AGAINST) HOLDING CORPORATIONS RESPONSIBLE ARE NOT RELEVANT TO AI AGENT RESPONSIBILITY	833
A. <i>Moral Reasons Theorists Offer for Holding Corporate Agents Responsible</i>	834
1. The Responsibility Deficits Reason	834
2. The Perverse Incentive Reason	835
3. The Quasi-Deterrent Reason	835
B. <i>Moral Reasons Theorists Have Offered Against Holding Corporate Agents Responsible</i>	837
1. The Harming Innocents Objection	837
2. Individual Wrongdoers are Off the Hook.	837
3. No Deterrent to Future Wrongdoings	838

* Assistant Professor of Management & Entrepreneurship, Leavey School of Business, Santa Clara University.

** The Charles J. Dirksen Professor of Business Ethics, Department of Management & Entrepreneurship, Leavey School of Business, Santa Clara University. © 2019, Vikram R. Bhargava and Manuel Velasquez.

4. Temptation to Prioritize the Group 839

C. *The Significance of Human Agents as Constituents of the Corporation.* 839

II. A MORE FUNDAMENTAL OBJECTION TO THE ANALOGY BETWEEN CORPORATE RESPONSIBILITY AND AI AGENT RESPONSIBILITY 840

A. *If the Human Members of a Corporation who Brought About a Corporate Act Are Morally Responsible for that Act, then the Corporation Would Be Excused.* 841

B. *If None of the Humans who Brought About a Corporate Act Are Morally Responsible for that Act, We Still Cannot Conclude that the Corporation Is Morally Responsible for that Act* 843

1. Intentional Attitudes and Attributions of Moral Responsibility to Corporations 846

2. Excuses and Attributions of Moral Responsibility to Corporations. 847

III. REFLECTIONS, PRACTICAL IMPLICATIONS, AND CONCLUSION. 849

INTRODUCTION

Several defenders of the claim that robots can be held responsible have appealed to the supposed reality of corporate responsibility to justify robot responsibility. Consider this collection of excerpts from theorists who draw attention to the purported fact of corporate responsibility and punishment as support for their claim that robots can be held responsible and punished for their actions:

Lawrence Solum writes: “The problem of punishment is not unique to artificial intelligences, however. Corporations are recognized as legal persons and are subject to criminal liability despite the fact that they are not human beings.”¹

Samir Chopra and Lawrence White state: “Finally, as for the educative function of punishment, while punishment of an artificial agent might not be educative for humans, it would nevertheless be educative for other artificial agents, given sufficient intelligence. After all, examples of corporate punishment are taken very seriously by other corporations.”²

1. Lawrence B. Solum, *Legal Personhood for Artificial Intelligence*, 70 N.C. L. REV. 1231, 1248 (1992).

2. SAMIR CHOPRA & LAURENCE F. WHITE, A LEGAL THEORY FOR AUTONOMOUS ARTIFICIAL AGENTS 169 (2011).

Gabriel Hallevy argues: “Criminal law recognized decades ago that the corporation, which is not a human entity, possesses life, freedom, body, and property. . . . But if the legal question concerning corporations, which are abstract creatures, has been decided affirmatively, it would be unreasonable to decide otherwise in the case of [artificial intelligence] systems, which physically simulate these human attributes much better than do abstract corporations.”³

Michael Anderson and Susan Anderson, describing the work of J. Storrs Hall, state: “[W]e have rules concerning corporations, to which robots of the future might be compared.”⁴

Peter Asaro states: “This will bring us to consider the punishments against other kinds of nonhuman legal agents, namely corporations, and what can be learned about robot punishments from corporate punishments. . . . Clearly, robots are different in many important respects from corporations. However, there are also many important similarities, and it is no coincidence that Coffee’s (1981) seminal paper on corporate punishment draws heavily on Simon’s (1947) work on organizational behavior and decision making, and in particular how corporate punishment could influence organizational decision making through deterrence.”⁵

It seems clear that these scholars think that the theoretical questions pertaining to corporate responsibility have at least something to tell us about robot responsibility. But as Peter Asaro aptly notes: “[A] great deal of work needs to be done in order to judge just how fruitful this analogy is.”⁶ This is the work we begin to take on in this paper.⁷

3. GABRIEL HALLEVY, *WHEN ROBOTS KILL: ARTIFICIAL INTELLIGENCE UNDER CRIMINAL LAW* 142 (2013).

4. MACHINE ETHICS 11 (Michael Anderson & Susan Leigh Anderson eds., 2011).

5. Peter M. Asaro, *A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, in *ROBOT ETHICS: THE ETHICAL AND SOCIAL IMPLICATIONS OF ROBOTICS INTELLIGENT ROBOTICS* 169, 170, 182 (Patrick Lin et al. eds., 2011).

6. *Id.* at 182. While it is of course correct that many of the above AI theorists are concerned with the question of whether non-human agents can be held legally responsible, that’s hardly worth investigating: there is of course legal precedent of holding non-human agent’s legally responsible. Moreover, these theorists seem to suggest that the metaphysical and moral grounds for attributions of responsibility of corporations might be helpfully considered in the context of AI responsibility. In any case, whether this is so is what we are concerned with. As Manuel Velasquez notes in the context of corporate moral responsibility, “The debate over whether corporate organizations are morally responsible for their actions is an attempt to influence and guide these prelegal understandings and, therefore, to indirectly influence and shape the laws that we subsequently craft. If the law’s treatment of corporate liability matters at all, then, it matters even more that our prelegal understanding of moral responsibility be correctly based because the law’s treatment of liability derives at least in part from these prelegal understandings.” Manuel Velasquez, *Debunking Corporate Moral Responsibility*, 13 *BUS. ETHICS Q.* 531, 537 (2003). Something similar seems to apply in the context of AI responsibility.

7. One scholar who has written on the topic of this analogy is Alan E. Singer. See Alan E. Singer, *Corporate Moral Agency and Artificial Intelligence*, 3 *INT. J. SOC. ORGANIZATIONAL DYNAMICS IN IT* 1, 1 (2003). He provides an account of how corporations and AI relate to one another from the philosophical pragmatist tradition.

In Section I of this paper, we explore the moral (as distinct from legal) reasons theorists have offered for and against holding corporations morally responsible.⁸ Holding an agent responsible goes above and beyond the mere belief or description that the agent is morally responsible or blameworthy. Holding an agent responsible involves—in addition to the belief that the agent is morally responsible—subjecting the agent to some negative response such as a hostile emotion (e.g., resentment and indignation), adverse social treatment (e.g., shunning or isolating), informal or formal sanctions (e.g., fines or firings), or punishment (e.g., imprisonment).⁹ We argue that each of the moral reasons theorists offer for holding corporations responsible are either inherently problematic or inapplicable to agents that possess artificial intelligence (hereinafter “AI agents”).¹⁰ For this reason, we conclude that the moral accounts that have been proffered for (and against) holding corporations responsible are not instructive for questions pertaining to the responsibility of AI agents.

Still, even if the moral reasons for and against holding corporations responsible are irrelevant in the AI context, perhaps the metaphysical accounts of corporate agency remain relevant. These metaphysical accounts take on the question of whether corporations are the sorts of entities that satisfy the conditions associated with moral agency. In other words, they try to show that corporations have certain metaphysical features that make them—like human persons but unlike boulders—the sorts of entities that *can* be morally responsible or blameworthy for some bad acts. Perhaps the metaphysical features theorists have constructed to show that corporations are moral agents that can be responsible for certain acts (putting aside the moral reasons for and against holding them responsible for these acts) could nevertheless be used to elucidate how AI agents likewise can be morally responsible for their acts.

Unfortunately, we do not believe the metaphysical corporate moral agency accounts developed thus far are useful for this purpose either. In Section II of the article we develop a more fundamental objection to accounts developed by

8. Carson Young, *Putting the Law in Its Place: Business Ethics and the Assumption That Illegal Implies Unethical*, J. BUS. ETHICS (forthcoming).

9. R. JAY WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS (1994); R. Jay Wallace et al., *Dispassionate Opprobrium: On Blame and the Reactive Sentiments*, in REASONS AND RECOGNITION: ESSAYS ON THE PHILOSOPHY OF TM SCANLON (R. Jay Wallace et al. eds., 2011); T.M. SCANLON, MORAL DIMENSIONS: PERMISSIBILITY, MEANING, BLAME (2008); GEORGE SHER, IN PRAISE OF BLAME (2006); MICHAEL MCKENNA, CONVERSATION AND RESPONSIBILITY (2012); DERK PEREBOOM, FREE WILL, AGENCY AND MEANING IN LIFE (2014); DAVID SHOEMAKER, RESPONSIBILITY FROM THE MARGINS (2015); Vikram R. Bhargava, *Firm Responses to Mass Outrage: Technology, Blame, and Employment*, J. BUS. ETHICS (forthcoming); Pamela Hieronymi, *The Force and Fairness of Blame*, 18 PHIL. PERSP. 115, (2004); Angela Smith, *On Being Responsible and Holding Responsible*, 11 J. ETHICS 465, (2007); BLAME: ITS NATURE AND NORMS (D. Justin Coates & Neal A. Tognazzini eds., 2013).

10. The definition of AI agents is controversial. Here we do not enter into this controversy but provide this informal and intuitive characterization: for the purposes of this article an AI agent is a human-made device that can perceive its environment and that can on its own make decisions about its actions in that environment, decisions that we could characterize as exhibiting some degree of intelligence or rationality.

defenders of corporate moral responsibility (hereinafter “CMR defenders”).¹¹ We argue that even if the theories CMR defenders have developed to show that corporations have the capacities to be held morally responsible are correct, these theories are insufficient to license judgments that hold that corporations are blameworthy for particular bad acts or outcomes. We suggest that this follows from the fact that corporations are made up of agents and act through agents. We then argue that since AI entities are not made up of agents nor act through agents, it may be easier to argue directly that AI agents are responsible, rather than trying to first demonstrate that corporations are responsible and then trying to argue from that demonstration that AI agents are responsible.

We conclude the article by briefly touching on what our arguments may imply for a recent European Union proposal to attribute legal personhood to smart robots. We argue that such proposals might be justified by pointing to the legal precedent of courts regarding corporations as legal persons, but appealing to corporations does not show us anything further about the responsibility of smart robots. In other words, while corporate legal personhood might serve as a sufficient precedent for granting legal personhood to non-humans, this fact does not bear on questions of whether robots themselves have the capacities required to be morally responsible for what they do.

If we are correct about the unfruitfulness of the analogy between corporations and robots, then the robot ethics scholars who argue for robot responsibility by appealing to corporations are making a mistake. We argue that whatever the truth of robot moral responsibility, it is not one we can establish by appealing to the purported responsibility of corporations.

I. THE MORAL REASONS FOR (AND AGAINST) HOLDING CORPORATIONS RESPONSIBLE ARE NOT RELEVANT TO AI AGENT RESPONSIBILITY

The argument we make in this section takes the form of a simple modus ponens with a conditional major premise: If neither the moral reasons for holding corporations responsible nor the moral reasons against holding corporations responsible make much sense when they are applied to AI agents, then this is a strong indication that, whatever the truth of corporate responsibility, corporations have little relevance for questions pertaining to AI moral responsibility. We accept this conditional premise as plausible without argument. So, the task will be to argue for the antecedent. For each of the reasons we consider, we will argue either that they do not make sense when applied to AI agents or that there are good independent grounds to doubt that they should be taken seriously (in either the AI context or in the corporate context). We begin by discussing the arguments CMR defenders have given to support the claim that corporations can be held morally

11. For influential defenses of corporate moral responsibility, see Peter A. French, *The Corporation as a Moral Person*, 16 AM. PHIL. Q. 207, (1979); CHRISTIAN LIST & PHILLIP PETTIT, GROUP AGENCY: THE POSSIBILITY, DESIGN AND STATUS OF GROUP AGENTS (2011); Philip Pettit, *Responsibility Incorporated*, 117 ETHICS 171, (2007) [hereinafter Pettit, *Responsibility*].

responsible, and then turn to the arguments that have been given to support the claim that corporations cannot be held morally responsible.

A. Moral Reasons Theorists Offer for Holding Corporate Agents Responsible

1. The Responsibility Deficits Reason

Some CMR defenders have argued that we have good moral reason to hold corporations responsible when there is a shortfall of individual responsibility for some bad outcome. Philip Pettit notes, “There is always likely to be a shortfall in enactor responsibility, and it is important to guard against that possibility,”¹² in part because individuals will often be “blamelessly ignorant of how their contributions will affect corporate performance.”¹³ Philip Pettit and Christian List also worry that if we do not take corporate moral responsibility seriously, this will mean “allowing some responsible actions . . . to go undetected.”¹⁴ So, they conclude, when there are no individuals who are morally responsible for a bad act done by a group, this results in a responsibility deficit, which could be narrowed or eliminated by holding the corporation responsible, given that corporations are supposedly capable of being held responsible.

The responsibility deficit line of argument for holding corporations morally responsible is mistaken. Pettit and List state that failing to hold corporations morally responsible when a responsibility deficit occurs “would mean allowing some responsible actions . . . to go undetected.”¹⁵ But actions are not responsible—*agents* are. Furthermore, we cannot presuppose that there must be some amount of responsibility that must be disbursed whenever some bad event or outcome occurs,¹⁶ even though it appears this is precisely what Pettit and List are assuming. They infer from the mere presence of what appears to be a bad event or outcome that there must be some amount of responsibility meted out to some agent or other.

Bad acts and bad events do occur, of course. But it is a separate question whether anyone was morally responsible for them. The occurrence of a bad act doesn’t automatically generate a particular amount of responsibility that must then be distributed to some agent or other. The very notion of a “shortfall of enactor responsibility” is a confused one.

As such, we have independent reason to dismiss the moral deficit argument for holding corporations responsible. Given that we have dismissed the moral deficit argument for corporate responsibility as based on a confusion, there is no reason to consider its applicability in the AI context.

12. Pettit, *Responsibility*, *supra* note 11, at 196.

13. *Id.*

14. LIST & PETTIT, *supra* note 11, at 166.

15. *Id.*

16. Matthew Caulfield and William S. Laufer call this mode of argument “topsy-turvy.” See Matthew Caulfield and William S. Laufer, *Corporate Moral Agency at the Convenience of Ethics and Law*, 17 GEO. J.L. & PUB. POL’Y (SPECIAL ISSUE) 953 (2019).

2. The Perverse Incentive Reason

The second moral reason CMR defenders have advanced for holding corporations responsible also relies on the notion of responsibility deficits, but nevertheless provides a distinct argument for holding corporations responsible. This is the argument that failing to hold corporations responsible would generate a perverse incentive for individuals to incorporate in order “to achieve a certain bad and self-serving effect, while arranging things so that none of them can be held fully responsible for what is done . . . [and so that they] are protected by excusing or exonerating considerations.”¹⁷ Moreover, Philip Pettit goes on to note, “Let group agents be freed from the burden of being held responsible, and the door will open to abuses: there will be cases where no one is held responsible for actions that are manifestly matters of agential responsibility.”¹⁸ In other words, if we do not hold corporations responsible, individuals could incorporate and act badly through the corporation in such a way that none of the individuals would reach the threshold for individual moral responsibility, and thus these individuals with untoward aims will be let off scot-free. Pettit thinks that failing to hold corporations responsible will create a perverse incentive to incorporate for those who are inclined to misbehave.

But this argument also rests on a confusion. The individuals who incorporate, of course, would not be free from moral responsibility—they would be morally responsible for intentionally creating an organization in furtherance of their untoward ends. The argument that not holding corporations responsible will generate perverse incentives to incorporate doesn’t get off the ground once we recognize that individuals *are* morally responsible for creating organizations with the aim of achieving bad ends (and also for then attempting to thwart and deceive others by trying to get others to regard them as not morally responsible). Accepting the claim that corporations are not themselves responsible for their actions will not create a risk that their members will be insulated from responsibility. Nor will the claim exonerate those who incorporate with the aim of shielding themselves from responsibility for some bad act they want to perpetrate through the corporation.

This second reason for holding corporations responsible can be dismissed outright since it too is based on a confusion. As such, we need not explore the relevance of this reason in the AI context either.

3. The Quasi-Deterrent Reason

The third moral reason CMR defenders give for holding corporations responsible (one that some think applies even to unincorporated collectives) is best characterized as a quasi-deterrent rationale.¹⁹ Defenders of this rationale argue that holding a group responsible will incentivize members of the group to challenge

17. Pettit, *Responsibility*, *supra* note 11, at 196.

18. *Id.* at 113.

19. *See generally* Pettit, *Responsibility*, *supra* note 11.

what the group does and will encourage the members to change the group constitution to avoid future bad acts. In support of this thought Philip Pettit states:

Think of the rationale for finding a commercial corporation responsible as a whole for some misdeed, rather than just finding the board or management responsible. Doing so is likely to provide an incentive for shareholders in that corporation, or in any similar corporation, to establish checks on the board and on the management.²⁰

Moreover, it might have a deterrent effect on other groups of a similar sort: “By finding the grouping responsible . . . we make clear to members of other groupings in the same category that they too are liable to be found guilty in parallel cases, should the body to which they belong bring about one or another ill.”²¹

This argument does not rest on a conceptual confusion or an oversight about where to locate individual responsibility. Nevertheless, it can be dismissed for two reasons. First, there is a strong moral objection as raised by John Hasnas.²² He notes that we can grant all of the purported deterrent effects associated with this rationale for holding corporations responsible, but it is nevertheless morally objectionable:

Fear of corporate punishment can motivate managers to institute compliance programs and make efforts to maintain a good corporate ethos that can reduce wrongdoing by employees. I have no doubt that this is correct. The threat of collective punishment is indeed an effective way to motivate people to suppress undesirable conduct by others. That is almost always its purpose. Collective punishment can deter. The problem is not that collective punishment is not effective. It’s that it is unjust.²³

That is, it might be quite right, as Pettit speculates, that holding an entire group responsible for the bad acts brought about by some of its members will motivate the other individual members to behave better or motivate them to improve the constitution of the group. Nevertheless, holding an entire group—including those who did not participate in the bad act—responsible for the acts of some of the members of the group is still unjust collective punishment.

Putting aside this objection to the quasi-deterrent rationale for holding corporations responsible, there is also good reason to think that this rationale is not sensible in the context of AI agents. Unlike corporations, AI agents are not constituted of human agents, so punishing an AI agent will have no deterrent effects on its constituents. While the AI entity might have been created by a human engineer,

20. *Id.* at 116.

21. *Id.* at 116.

22. John Hasnas, *Reflections on Corporate Moral Responsibility and the Problem Solving Technique of Alexander the Great*, 17 J. BUS. ETHICS 183 (2012).

23. *Id.* at 192.

holding the AI entity responsible will do little to deter the engineer's bad behavior nor will it develop the engineer's moral outlook. Of course, criticizing or disparaging an AI agent may motivate its maker to change the agent's design. But criticizing the operation of a device is quite different from holding the device morally responsible.

We can conclude, then, that the three moral reasons (the responsibility deficit, the perverse incentive, and the quasi-deterrent reasons) proffered by scholars for why we should hold corporations responsible have little relevance for AI agents. We can dismiss those reasons as mistaken on independent grounds or as not applicable in the AI context. We now turn to the moral reasons critics of corporate moral responsibility have offered against holding corporations morally responsible. Again, we will argue that these moral reasons have little relevance to questions of AI responsibility.

B. Moral Reasons Theorists Have Offered Against Holding Corporate Agents Responsible

1. The Harming Innocents Objection

One significant moral consideration against holding corporate agents responsible is that the harms associated with blame and criminal prosecution will fall upon not merely the individuals who acted wrongly in the corporation, but also those who are entirely innocent.²⁴ It is highly implausible, for example, that the custodial staff of a corporation have much to do with the corporate wrongdoing. As Velasquez notes: “[T]he decision to hold the organization as such responsible for the crime, and the related decision to punish the organization as such, [results] in the innocent being forced to suffer along with the guilty.”²⁵ The costs of the punishment will be passed on to stakeholders who are innocent of wrongdoing such as employees, consumers, or shareholders.²⁶ This objection is a plausible one against holding corporations responsible.

However, this sort of objection does not make much sense in the context of punishing AI agents because AI agents are not groups of people. There is no potential human agent we might innocently punish by holding an AI agent responsible. As such, this particular moral reason against holding corporations responsible is not relevant in the AI context.

2. Individual Wrongdoers are Off the Hook

Velasquez also objects that if we start regarding corporations as responsible, then it is possible that “the individuals through whom the corporation acted—that is, the individuals who actually brought about the corporate act—are never

24. *Id.* at 191.

25. Velasquez, *supra* note 6 at 536.

26. Hasnas, *supra* note 22, at 191.

brought to justice and never given a just punishment.”²⁷ This objection against holding corporations responsible is also plausible. More importantly, this objection *is* potentially importable in the context of AI responsibility. The thought might go: if we begin focusing on holding AI entities responsible, the individual agents involved in programming the AI might be let off the hook.

But note that this concern in the context of AI agents is not analogous to the concern in the corporate context. The concern with holding corporations responsible is not that the individuals would escape moral responsibility for starting the corporation, or for designing the corporation in a certain way. Rather, the concern is that individuals would escape individual moral responsibility for a particular bad corporate act. The question of moral responsibility for starting a particular type of organization is a different one, and is decidedly a human responsibility issue. The issue arises in the corporate context because corporate acts are done through human agents. But there *is* no parallel issue in the context of AI, given that AI acts are not done through humans.

So, the relevant question is whether holding the AI entity *itself* responsible for X-ing would allow individuals to escape responsibility for X-ing. But X-ing is not done through human agents in the AI context; it is done through the AI agent itself. In short, this line of thought, while plausibly a sound objection to holding corporations responsible, does not make sense in the AI context because the AI entity does not act through human agents. It acts by itself and through itself.

3. No Deterrent to Future Wrongdoings

Another objection against holding corporations responsible is that in instances where we punish the corporation, this punishment can “fail to touch the individuals who in reality carried out the crime and thus fail to control effectively the socially injurious behavior of the corporate organization.”²⁸ While this moral reason may appear similar to the reason discussed above, it is not quite the same. The previous reason pertained to individual wrongdoers escaping accountability for some past wrongdoing. This reason pertains to the worry that devoting our attention to holding corporations responsible would leave little deterrence in place against individual wrongdoers perpetrating or contributing to future bad acts in the group context. That is, once we start targeting groups for punishment, this creates a deterrence void with respect to individuals within that organization.

This objection might appear to have relevance for the AI agent context at first blush. But whose wrongdoing are we seeking to deter in the AI context? If we are seeking to deter an individual programmer’s wrongdoing, then we need not appeal to corporations at all to address that issue. Our existing individual moral responsibility theories can handle that. On the other hand, if we are seeking to deter the human agents who were somehow involved in the AI agent’s action, then it is a non-issue: no human agents were involved since AI agents act

27. Velasquez, *supra* note 25, at 537.

28. *Id.* at 536.

autonomously. AI agents are not operated by humans in any meaningful sense but are designed to act on their own.

So, again, the fact that the AI agent is not acting through a human agent or constituted of agents becomes relevant. There is no human agent to deter from performing a particular AI agent act because human agents did not contribute to performing that act—even if human agents contributed to creating the AI agent.

4. Temptation to Prioritize the Group

The last objection to holding corporations morally responsible we consider is this: insofar as we begin to treat corporate agents as morally responsible entities, we risk prioritizing corporate interests ahead of the interests of individuals. Velasquez notes:

We will be tempted . . . to look upon the corporation as organic theories of the state looked upon the state: since the corporation is a whole person (with its own group mind) and the member merely a part, the interests of the corporation's members may legitimately be sacrificed to the corporation's interests and the good of the individual may be subordinated to the corporation's good.²⁹

The thought is that because the corporation is a sort of supra entity, considerations of its constituent parts (human agents) may be seen as less morally important. This is a dangerous possibility.

While this objection might again be a sound one in the context of objections to corporate moral responsibility, it does not make sense in the context of AI agents. This is because AI agents are not made up of sub-agents. There *is* no worry of subordinating a particular motherboard or computer chip to the AI agents given that these subparts are simply not agents.

C. The Significance of Human Agents as Constituents of the Corporation

The lesson taken from the discussion thus far should be that several of the defenses and objections to holding corporations responsible derive their significance, at least in part, from the fact that corporations are made up of human agents and act through human agents. Nearly all of the reasons we have considered regarding whether to hold or not hold corporations responsible appeal to the implications of corporate responsibility for the human agents which make up the corporation. But AI agents, though perhaps initially programmed by human agents, are not *made up* of human agents or any other agents, nor do they need to act through human agents or other agents.

29. Manuel Velasquez, *Why Corporations Are Not Morally Responsible for Anything They Do*, 3 BUS. & PROF. ETHICS J. 1, 15 (1983).

For example, consider an AI entity whose algorithms repeatedly yield certain racist patterns in the mortgage loans it approves.³⁰ Suppose at the time of programming, the AI agent was programmed by a racist programmer. But later the racist programmer comes to see his moral errors and abandons his racist beliefs. This won't suddenly and automatically change the AI agent. The AI loan approval agent still has these racist algorithms and still continues its racist operations. A similar point cannot be made for groups such as corporations. Consider a group that seems to be acting in ways that are racist. But then suppose all members of the group abandon their racist beliefs. The group, too, would then be fundamentally altered and would not continue its racist behavior. Insofar as the members of the group reformed their racist beliefs, they would also begin to remove any racist group policies or traditions.

The point is simple: once programmed, the AI is no longer reliant on the programmer. It interacts with the world and so on without the need for the programmer to serve as the puppeteer. But the corporation requires the human persons that constitute it to act—neither side of the corporate moral responsibility debate disputes this. Human persons play a critical role in corporate action. Humans do not play such a role in AI agent action. This is so, even if we admit that the genesis of both the corporate agent and the AI agent are due to a human agent. The fact that there are human agents that constitute the corporation and that corporate acts occur through those human agents is critical to our moral reasons for—and against—holding corporations morally responsible. But AI agents are not made up of agents. Thus, these moral reasons are not relevant in the AI context.

II. A MORE FUNDAMENTAL OBJECTION TO THE ANALOGY BETWEEN CORPORATE RESPONSIBILITY AND AI AGENT RESPONSIBILITY

In the previous section of our article (Section I), we evaluated how relevant the moral reasons for and against holding corporations responsible might be to questions of AI agent responsibility. We argued that none of these specific reasons have import for questions pertaining to AI responsibility. Most of the moral reasons we examined—both for and against holding corporations responsible—fail to shed any light on questions related to AI responsibility. They fail either because those reasons can be rejected outright as unsound, or because the acts of AI agents are not done through human agents and AI agents are not composed of human agents.

One might object that while it may be true that the moral reasons why we should (or should not) hold corporations responsible are not relevant in the AI context, the metaphysical accounts that purport to show how corporations can properly be regarded as moral agents may still be relevant in the AI context. There may, after all, be other moral reasons to hold AI agents morally responsible, entirely different from those we criticized in the previous section. As a result, accounts of corporate moral agency, the objection might go, could still be

30. Nick Bostrom & Elizer Yudkowsky, *The Ethics of Artificial Intelligence*, in CAMBRIDGE HANDBOOK OF ARTIFICIAL INTELLIGENCE (William Ramsey ed., 2011).

instructive for understanding AI agent responsibility, even if the moral reasons for holding AI agents responsible are not supported by the moral reasons for holding corporations responsible.

To deal with this objection we will now try to show that there is something more fundamental to the nature of corporate moral responsibility and corporate agency that precludes corporations from, even in principle, being helpful for understanding AI responsibility. Specifically, the fact that corporations are composed of agents and must act through those agents while AI agents are neither composed of other agents nor do they act through other agents is significant for (at least) two further reasons. First, the fact that corporate actions require acting through humans is reason to believe that even if corporations commit a bad act, their wrongdoing will nearly always be excused. Second, the fact that corporations act through human agents will undermine our ability to ascribe responsibility to the group when there is a supposed “responsibility deficit” (a concept we earlier suggested was problematic but which, for the sake of argument, we will assume makes sense). Neither of these two reasons apply in the AI context, we will show, because AI entities do not act through humans. This fundamental difference between corporate agents and AI agents, we argue, undermines the possibility that corporate responsibility can shed light on AI agent responsibility.

Going forward we will accept for the sake of argument that defenders of corporate moral responsibility are correct about corporations possessing the kinds of capacities (such as beliefs, intentions, etc.) that make them capable of being morally responsible for their acts. In doing so, we will put to one side the various arguments that critics of corporate moral responsibility have raised about why groups can only be said to have these capacities in a metaphorical or “as-if” sense, or as a way to speak more concisely of the particular sets of attitudes that the individuals constituting the group possess. Going forward, let’s suppose CMR defenders have succeeded in arguing for the view that groups *do* possess the needed capacities to be held responsible.

A brief outline of what is to come will help: That all corporate acts can only occur through humans is something neither side in the CMR debate denies. Given this, for any given bad act that appears to be a corporate act there are two possibilities we consider: (1) some or all individual human actors who brought about the act are morally responsible for the act and (2) none of the individual human actors who brought about the act are morally responsible for it (and in this latter case, theorists like Pettit argue a “responsibility deficit” arises). We suggest that in neither of these two possibilities will we be justified in judging that a corporation is blameworthy for the given bad act.

A. If the Human Members of a Corporation who Brought About a Corporate Act Are Morally Responsible for that Act, then the Corporation Would Be Excused

In this sub-section, we consider the first possibility: instances where some or all of the individuals are morally responsible for the group act. Suppose there is

some bad corporate act that is perpetrated through individual human agents. The Ford Pinto case might be considered a paradigmatic example of something like this. Let's further suppose that the individual human agents who were causally responsible for the bad act are also morally responsible for it. CMR defenders, in such a scenario, would hold that the relevant beliefs, desires, and intentions that render the individual human agents morally responsible, can properly be attributed to the group, and the group would then be morally responsible for the bad act. In short, the corporation *and* the individual's moral responsibility is grounded in the same beliefs, desires, and intentions.

But this presents a problem for CMR defenders who would hold that the corporation is morally responsible for this act. This is because the corporation did not have a choice in the matter and lacked control over whether it would perpetrate that act or not. As one scholar puts it, "Collective agents are too much at the mercy of their constitutive individual members for it to be fair to hold collective agents morally responsible in their own right."³¹ And some form of control is commonly thought to be a crucial component of ascriptions of moral responsibility.³² In other words, the corporation could not but do what its members demanded, given that the corporate act was perpetrated through the agents, and the corporation's beliefs, desires, and intentions were injected by the individual agents.

We are reminded of the familiar evil neuroscientist who operates on a person's brain in such a way as to make it impossible for that person to do anything but commit the bad act that the evil neuroscientist tweaked the person's brain to do. The corporation was made to do what its members insisted it do—given that it could only act through its members—and it had no meaningful way to reject its member's inputs, and so it could not plausibly be morally responsible for the act. The corporation could not be morally responsible for that bad act because its autonomy is so completely undermined that it would have to be excused for performing that bad act.

The corporation has an excuse in nearly all instances in which it is made to perform a bad act by its constituent agents: namely, that it was forced to act in the way its members directed. At best, we can say that the corporation performed this bad act and ought to experience some sort of agent regret. But this falls far short of moral responsibility for that bad act.

AI agents, of course, differ from corporate agents. AI agents need not rely on the actions of their programmers to act, in the same way that corporations must rely on the actions of their constituent members to act. Nor are AI agents made up of programmers in the way that a corporation is made up of its human agents. With any autonomous AI entity, this particular issue we have highlighted of corporations almost always being excused for any bad act or outcome due to their lacking autonomy over the decision, does not arise. So, there is at least one hurdle

31. Pekka Mäkelä, *Collective Agents and Moral Responsibility*, 38 J. SOC. PHIL. 456, 466 (2007).

32. *Id.* at 465.

that theorists of AI responsibility will not face. And given that corporate responsibility defenders face this hurdle, it is not sensible to appeal to a more problem laden view (namely corporate responsibility views) to defend AI responsibility.

B. If None of the Humans who Brought About a Corporate Act Are Morally Responsible for that Act, We Still Cannot Conclude that the Corporation Is Morally Responsible for that Act

What about the second possibility, one where there is some bad act/outcome that appears to be due to the corporation but for which no individual member of the corporation is morally responsible? This second possibility involves the sort of case that some may regard as lending the most intuitive support for the possibility of corporations being autonomous moral agents. It is the kind of case in which no member meets the threshold for individual responsibility, yet nevertheless it appears the corporation performed some bad act.

As mentioned earlier, theorists like Pettit believe such a situation generates a responsibility deficit and that the corporation should be held responsible to reduce or eliminate the deficit. In an earlier section, we argued that the idea of a responsibility deficit was problematic, given that responsibility attaches to agents and that bad acts do not create certain allotments of responsibility that must always be doled out to some agent or other. But, here, for the sake of argument we put that point to the side and suppose that the notion of a responsibility deficit is perfectly coherent. Recall also we are supposing for the sake of argument that CMR defenders have successfully argued for the view that corporations have all the relevant underlying capacities to be held responsible; we can even grant that they have the ability to have phenomenological states and reactive attitudes.³³ What might follow from this? Would such a theoretical accomplishment license what CMR defenders seek, namely the judgment that corporations are blameworthy for a given act in a way that makes them apt targets of blame or punishment?

We will argue that it does not. First, in the wake of some bad act/outcome, we need to sort out whether it was perpetrated by the corporation in the first place, before we can get to the question of moral responsibility for that act. It is not enough that it seems as though the corporation perpetrated the bad act. Moreover, we argue, even when corporations possess the correct interior psychological states, as it were, this does not license the claim that the corporation performed the bad act. In short, we will show that even if the arguments advanced by CMR defenders are sound, they have argued for a sort of corporation in a vat and their arguments do not justify the view that corporations are morally responsible for a particular bad act.

Consider something like the recent oil spill for which, some have claimed, BP was responsible. Let's grant that the oil spill is a bad outcome and that it is something that is, in Pettit's words, an act that is "manifestly [a matter] of agential

33. Gunnar Björnsson & Kendy Hess, *Corporate Crocodile Tears? On the Reactive Attitudes of Corporate Agents*, 94 PHIL. & PHENOMENOLOGICAL RES. 273 (2017).

responsibility,”³⁴ and not something like a bad outcome without an agential source such as an earthquake. Does the fact that BP is the sort of entity that possesses all the capacities associated with moral agency (by assumption) support the conclusion that BP (over and above the individual members of BP) is morally responsible for the oil spill when no individual member of BP is morally responsible? We maintain that it does not.

Suppose we are at the *MoMA* and someone has spray painted “ugly” on Piet Mondrian’s *Composition in Red Blue and Yellow*. Now the mere fact that I am a visitor at the museum and possess the capacities to be held morally responsible, does not mean that the act of defacing the painting can justifiably be attributed to me, nor does it justify the judgment that I am morally responsible for defacing the painting. This is so, even if I might be morally responsible for all sorts of other acts, including defacing other paintings. The same goes for attributing the act to the other guests at the museum (even if we know that at least one of the thousands of visitors currently at the museum must be morally responsible for defacing the painting). The point is a simple and obvious one: the occurrence of a bad outcome, even a bad outcome that clearly has an agential causal source, together with the presence of entities with the capacities to be held morally responsible, do not alone justify holding any specific one of the moral agents present morally responsible for the act.

Similarly, the mere fact that there was an oil spill and that corporations are capable of being held morally responsible does not, by itself, imply that a particular corporation was morally responsible for that spill. It is a further matter whether the corporation is responsible for the particular bad outcome in question (i.e., *this* oil spill). In the wake of some bad outcome (for which CMR defenders are tempted to hold the corporation responsible), we cannot assume that the act is properly attributed to the corporation (let alone that the corporation is morally responsible for that particular bad outcome) just because it was an entity with the proper capacities to be a candidate for blame. Other candidates for blame for the bad act include you, me, the barista, the flight-attendant, the emergency room nurse—any of us who are agents fit to be morally responsible, as well as *other* corporations (who by hypothesis are full-fledged moral agents). But the occurrence of some bad act or outcome, joined with the presence of proper candidates for blame, does not make those particular candidates for blame morally responsible for the bad act or outcome in question. The occurrence of a bad outcome for which it does not seem any individuals of a group are morally responsible does not license attributing that bad outcome to the group (and in turn does not license attributing moral responsibility to the group for that act), even if we grant that corporations have the capacities to be held morally responsible.

A CMR defender might concede this much and state, “Of course, merely possessing moral responsibility capacities is not enough; the content of the attitudes

34. Pettit, *Responsibility*, *supra* note 11, at 197.

that in part make up these capacities matters! Our accounts indicate how corporations can be said to have beliefs and intentions, and of course it matters to responsibility ascriptions for a particular bad outcome that the corporation had the relevant syndrome of beliefs and intentions pertaining to that bad outcome.” In other words, CMR defenders might say that they are not claiming that merely possessing agential capacities is sufficient grounds for being held responsible. Rather, they are claiming that corporations *do* possess agential capacities, and if we are trying to sort out whether a bad outcome can be attributed to the corporation and whether the corporation can be said to be responsible for that bad outcome, we ought to examine the intentions and attitudes the corporate agent possesses to see whether they are related to the bad outcome in question. If they are, then the corporation can properly be said to be morally responsible for that bad outcome.

This, too, however is inadequate grounds for attributing the bad act to the corporation (and in turn holding the corporation responsible for the bad outcome). Even if the corporation has the intention and desire for the bad outcome or wrongdoing, it still does not allow us to attribute the bad outcome to the corporation, nor hold it responsible for that bad outcome. Suppose I intend and desire to rob a store—but have not in fact robbed the store. Before I get to act on my intention, I show up and the store has already been robbed. But no robber is found. Now, just because they have not yet found any individuals morally responsible for the robbery, this doesn’t mean that the robbery can be attributed to me, even if I was cheering for it to happen (and would have robbed it myself, had it not already been robbed). At best, I might be held responsible for having untoward beliefs and intentions, insofar as one believes that agents can be held responsible for their attitudes.³⁵

It seems the CMR defender’s arguments, even if successful, simply amount to defending something akin to a corporation in a vat, so to speak. At best, the arguments, even when successful might allow for us to criticize certain untoward beliefs, desires, attitudes, and intentions a corporation possesses, but it does not allow us to attribute bad acts to the corporation and thus precludes us from holding the corporation responsible for those bad acts. As for the sanctions CMR defenders are after, even if they are right about the capacities and metaphysics of corporations, their accounts amount to providing justificatory support equivalent to justifying blaming a brain in a vat for possessing murderous attitudes, or holding a person responsible for wishing untoward things upon an enemy. But this would hardly amount to the brain in the vat *being* a murderer, or the person perpetrating those untoward things. CMR theorists are missing an account of how a given bad outcome links up to a corporation’s intentional action. It is not enough that whatever state of affairs the corporation intended to bring about is realized. Thus, even if CMR theorists are entirely right about the metaphysics of group

35. Angela M. Smith, *Responsibility for Attitudes: Activity and Passivity in Mental Life*, 115 *ETHICS* 236 (2005).

intentions and minds, this at best allows us to hold a corporation responsible for possessing untoward beliefs, desires, and intentions, but not for bad outcomes or acts. Their account would not allow us to say that a particular bad act is due to the corporation, nor, a fortiori, that the corporation is morally responsible for that bad outcome.

A more sophisticated CMR defender might still have a response available: “Of course the mere presence of intentional untoward attitudes do not ground moral responsibility for some bad outcome that the attitudes are about. Nor does the presence of agents in the wake of a bad outcome entail that any specific agent is responsible for that bad outcome. Moreover, it’s correct that in responsibility deficit cases, no individual member is morally responsible. But none of this means that the individual member is not *causally* responsible for bringing the bad act about. So, our view is that the members still brought about the corporate act, despite not being morally responsible for the act. Thus, in responsibility deficit situations, the group is morally responsible when the members are causally responsible for a bad act X *and* the group possesses the relevant intentional attitudes pertaining to X.”

However, this response of the CMR defender is still inadequate for two reasons. It is inadequate, first, because it fails to provide an adequate explanation of how the act for which the organized group is said to be responsible, was brought about by the group’s intentional attitudes. And, second, it fails to provide an adequate explanation of the role that excuses can play. We discuss these two problems in the next two sections.

1. Intentional Attitudes and Attributions of Moral Responsibility to Corporations

First, then, what is missing in the above response of the CMR defender is an account of how an act perpetrated by a member of an organization (supposedly a member who is not morally responsible for that act) was brought about by the group’s intentional attitudes. Consider the following example. Mafia Don:

I spend my days brooding about murdering Sollozo and intend to kill him the next time I encounter him. I don’t order any of my underlings to kill him because I want the pleasure of killing him. While we are at a large wedding, I see Sollozo and thoughts of murdering him come to mind; I plan to murder him as we transition to the reception. My toddler, whom I am carrying in my arms, thinking that he is playing with his toy gun (but as it turns out, it is a real, loaded gun, given to him by accident), shoots and kills Sollozo.

In this case, the toddler is causally responsible for the death of Sollozo, but is not morally responsible. There is, therefore, what Pettit would call a responsibility deficit. But is the Don then morally responsible for murdering Sollozo in this case? It seems clear that he is not. While the Mafia Don might very well be morally responsible for possessing morally untoward intentional attitudes, he is

not morally responsible for murder. One's intentional attitudes need to be causally efficacious with respect to the murder for one to be morally responsible for the murder.

We might say something similar in the context of judging that a corporation is blameworthy for a particular bad act or outcome. We cannot say that corporation C is morally responsible for X-ing, merely because members of C were causally responsible for X-ing and C possessed the relevant morally untoward intentional attitudes about X-ing. What is required is an account of how C's untoward intentional attitudes brought about the member's X-ing in a way that would render C morally responsible for X-ing.

Moreover, it's also not enough for C's untoward attitudes to merely be causally efficacious with respect to bringing about a member's X-ing—it has to be causally efficacious in a certain kind of way. For example, if I merely suggest to a person, or even beg or pressure that person to carry out some bad act, and that person then follows through and performs that bad act, that alone will not make me morally responsible for that bad act. For example, if I beg a friend to punch the rude barista and the friend in fact does so, it does not necessarily follow that I am morally responsible for punching the barista (though it might make me morally responsible for pressuring a friend to do that bad act). Indeed, even if I know that my friend is one who is quick to physically escalate things and has a hot temper, it does not make me morally responsible for the *punch*, even if it may make me morally responsible for manipulating a friend to satisfy my objectionable personal desires.

For me to be responsible for another person's bad act, I would need to stand in a special sort of causal relation to the person—I would need to be in a position to make the person do something (perhaps through coercion, drugs, hypnotism, brain manipulation, etc.). One might point out, that in the case of corporations, there *is* an authority relation that puts the corporation in a position to order a person to do things—after all, the corporation is paying the salary. So, if the corporation has the relevant beliefs, desires, and intentions for a bad outcome Y, and then orders a member of a corporation to bring about Y, and that member does bring about Y, then we might say that the corporation made the member act, and so is morally responsible for Y.

But this would be a troubling line to pursue for a CMR defender. First, the members of the corporation who order the bad act are the very ones whose beliefs, desires, and intentions are figuring into the corporation's decision to order a member to X. It does violence to the concept of "ordering someone to X" if the entity being ordered is also doing the ordering. This is because the orders that the group is giving in responsibility deficit situations are arrived at by way of individual member inputs, even if the order may not have been the most preferred choice of any individual member.

2. Excuses and Attributions of Moral Responsibility to Corporations

Second, the response of the CMR defender described above is also missing an explanation of the role that excuses can play in attributions of moral

responsibility to corporations. To the extent that we want to use our human concept of moral responsibility as a model for corporate responsibility (what *else* would we model it off of, after all)—and we want to hold that they are agents in much the way we, humans, are—it is imperative to consider responsibility mitigating dimensions too. If we want to admit group agency, we need to admit excuses. If corporations are not subject to excuses, then this would imply that corporations are automatically responsible for a bad act. It would be like holding a parent responsible for *anything bad* that happens to his or her child no matter the circumstance. On such a view, one could forcibly restrain me and then do bad things to my child and I'd still be morally responsible for the bad thing done to my child. This is clearly implausible.

Any account of group moral responsibility, then requires an account of the circumstances under which a group is *not* morally responsible for a bad act it performed. It is not just that theorists have not addressed the issue of group excuses, rather, the trouble is that current accounts of corporate responsibility leave no room for group excuses. Corporate responsibility theorists have collapsed the question of whether a corporation is morally responsible for an act (in the sense required for blame and punishment) into the question of whether a bad act can properly be attributed to a corporation. But answering the latter question affirmatively does not license blame and punishment.

Moreover, CMR defender's accounts do not leave room for us to be *mistaken* about moral responsibility ascriptions of groups. If no individuals are morally responsible and there is some bad outcome the corporation produced, then the corporation, CMR defenders would have to conclude, must be morally responsible for that bad outcome. Let's suppose it turns out a foreign government official had a mole planted in BP who was responsible for the oil spill. It turns out that it had nothing to do with BP's policies. Would it no longer be the case that BP is morally responsible? It seems clear that BP is not morally responsible. *But what were we mistaken about?* CMR defenders' accounts lack grounds to answer that question, since it is by fiat taken that the group is responsible for a bad group act when no individuals were responsible. Did the group suddenly stop being responsible because we later found an individual to hold responsible? This would be implausible, given that little about our factual understanding of the group's properties would have changed.

The challenge for corporate moral responsibility theorists is this: If theorists cannot answer what it would look like for a firm to have performed a bad act, but still not be responsible for that act, then the account is fundamentally inadequate. We can all perform lots of bad actions for which we are not responsible. We can be tripped, bumped, forced, coerced, and duped into performing bad acts. But we are not morally responsible for the act under these circumstances.

One might object that there *are* such examples that can be constructed. For instance, suppose a member of the group is held hostage and the group is told that unless the group performs some bad act X, then the member will be killed. So, the members convene to vote on performing X and ultimately the group performs

X under the threat of one of its members being killed. In such a case, insofar as X is some bad act, the group might be excused for X-ing.³⁶

Let's grant that this is indeed an example of a group being excused for performing some bad act. Unfortunately, the above example is one in which the group's attitudes and intentions will at least partially, if not entirely, be captured by the member's intentions and attitudes. But CMR defenders acquire much of their intuitive support for the claim that corporations can act autonomously from discursive dilemma situations in which a group act is not fully accounted for by any individual member's desires, beliefs, or intentions. CMR defenders, then, must articulate a case in which a group performs some bad act for which it was excused, where the attitudes and intentions underlying the bad act are not also entirely captured by the members' attitudes and intentions. Such an example does not come to mind easily.

Our point is simple: the claim that corporations have the capacities to be held morally responsible, even if true, provides insufficient grounds to hold corporations morally responsible for a given wrongdoing or bad outcome. Moreover, even if this problem is surmountable, any plausible account of moral responsibility must be able to articulate the circumstances under which an agent is not morally responsible or is excused for an outcome its actions produced—otherwise, the agent will be taken to be morally responsible by fiat for any bad outcome that its actions produce.

What does this mean for theorizing about AI agents? It means that accounts developed of corporate moral agency lack the theoretical resources to ground AI agent responsibility and that whatever the truth of AI agent responsibility, it is not one to be found by examining accounts of corporate moral responsibility. Given that corporations are made of humans and can only act through humans, theories of corporate moral responsibility have a bug that theories of AI responsibility will not have.

III. REFLECTIONS, PRACTICAL IMPLICATIONS, AND CONCLUSION

It would be peculiar, dare we say unfortunate, if scholars began turning their attention to the questions of whether toothbrushes and trash bins, or even entities with more complicated designs, like the Taj Mahal or a thermostat, had the capacities to be held morally responsible. Why? Whether or not these things are (or are not) morally responsible seems to matter only to the extent that it bears on our moral responsibility practices. Given that we rarely are tempted to subject these above items to sanctions, concerns about whether they have the capacity to be held morally responsible do not seem worth pursuing. But unlike with toothbrushes and thermostats, many are tempted to hold corporations morally responsible—over and above the individual constituents of the corporation—for wrongdoing by way of sanctions, typically in the form of blame or punishment.

36. We thank David Shoemaker for this example.

And it seems, by looking to the example of corporations, some robot ethicists are attempting to justify accounts of AI entity responsibility for similar purposes.

The trouble, however, is that most recognize that the mere temptation to blame and punish is not enough to justify blame and punishment. This is because one must be an agent in order to be a proper recipient of blame and punishment. For this reason, CMR theorists have developed sophisticated accounts of how corporations, over and above their constituent member's moral responsibility, have the capacities to be held morally responsible for bad acts and can thus justifiably be subjected to the sanctions associated with our moral responsibility practices, including blame and punishment. But we have argued that neither the moral reasons theorists have offered for holding corporations morally responsible, nor the theoretical underpinnings of corporate moral agency, are helpful with respect to questions of AI responsibility.³⁷ As a result, the question of whether or not we can and should hold AI entities responsible remains an open one.

But the "what's the point of holding it responsible?" question emerges even more forcefully in the case of holding AI agents responsible. Suppose an autonomous vehicle crashes in a way that kills a person.³⁸ What would it mean to hold this autonomous vehicle morally responsible for that death? Should we program a virus into the autonomous vehicle? That seems about as productive as shooting arrows into the sky to hold the sky responsible for the badness associated with a tornado. This brings out at least one salient fact: even if suffering is not the *aim* of blaming and holding responsible, that it typically plays an important role is undeniable ("the sting it putatively ought to have when directed at one who is blameworthy"³⁹). When punishing corporations, in fact, *somebody* feels the sting (usually the shareholders). This, however, is not so with current versions of AI agents. Perhaps some future AI agents will have the capacity to suffer and to feel this sting, but we are not at that point yet. And even when we are at that point, we still will not improve our understanding of advanced AI agent responsibility by appealing to corporate responsibility.

We close with a practical implication: The European Union (EU) has recently begun dealing with questions pertaining to the moral status of robots.⁴⁰ If we are correct, the theories that have been elaborated to justify attributing moral responsibility to corporations cannot help the EU determine whether robots should be held morally responsible for their acts. If, on the other hand, the question is

37. There is of course important work to be done on broader issues at the intersection of artificial intelligence ethics and business ethics. For some recent work in the area, see Tae Wan Kim & Andrew Scheller-Wolf, *Technological Unemployment, Meaning in Life, Purpose of Business, and the Future of Stakeholders*, J. BUS. ETHICS (forthcoming 2019).

38. Vikram Bhargava & Tae Wan Kim, *Autonomous Vehicles and Moral Uncertainty*, in ROBOT ETHICS 2.0: FROM AUTONOMOUS CARS TO ARTIFICIAL INTELLIGENCE 5, 5 (Patrick Lin, Keith Abney, & Ryan Jenkins ed., 2017).

39. MCKENNA, *supra* note 9, at 121.

40. James Vincent, *Giving Robots 'Personhood' Is Actually about Making Corporations Accountable*, THE VERGE (Jan. 19, 2017, 10:45 AM), <https://www.theverge.com/2017/1/19/14322334/robot-electronic-persons-eu-report-liability-civil-suits> [<https://perma.cc/PA8P-47KQ>].

simply whether there is legal precedent for treating robots as if they are responsible, the answer must of course be answered affirmatively. The law may, by fiat, treat not just corporations, but any kind of entity as if it is responsible (the law can, and historically has, attributed criminal responsibility to pigs, dogs, grasshoppers, beetles, rats, flies, snails, worms, and various pests).⁴¹ But such legal attributions of responsibility to corporations or other entities shed little light on the nature of moral responsibility nor on whether robots are the kind of agents that can be morally responsible for their acts. Insofar as robot ethicists are interested in any of the theoretical import from corporate responsibility, there won't be much to be found. As a result, AI responsibility accounts might be able to appeal to the legal precedent of holding corporations responsible. But beyond this legal point, appeals to corporate responsibility can provide no real insight into what moral responsibility might mean when attributed to AI agents.

41. EDWARD P. EVANS, *THE CRIMINAL PROSECUTION AND CAPITAL PUNISHMENT OF ANIMALS*, (Farber & Farber ed. 1987).