View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DigitalCommons@Macalester College

Macalester College
DigitalCommons@Macalester College

Philosophy Honors Projects | Philosophy Department

Spring 5-2-2018

Animals, Machines, and Moral Responsibility in a Built Environment

Logan Stapleton
Macalester College, lstaple99@gmail.com

Follow this and additional works at: https://digitalcommons.macalester.edu/phil_honors

 Part of the Philosophy Commons

Recommended Citation
Stapleton, Logan, "Animals, Machines, and Moral Responsibility in a Built Environment" (2018). *Philosophy Honors Projects*. 12.
https://digitalcommons.macalester.edu/phil_honors/12

This Honors Project is brought to you for free and open access by the Philosophy Department at DigitalCommons@Macalester College. It has been accepted for inclusion in Philosophy Honors Projects by an authorized administrator of DigitalCommons@Macalester College. For more information, please contact scholarpub@macalester.edu.

Animals, Machines, and Moral Responsibility in a Built Environment

Logan Stapleton
Macalester College Philosophy
Advisor: Diane Michelfelder

Abstract: Nature has ended. Acid rain and global warming leave no place untouched by human hands. We can no longer think of 'the environment' as synonymous with 'nature'. Instead, Steven Vogel argues that the environment is more like a mall: it is built. And because we build the environment, we are responsible for it. Yet, other things build, too. Animals build and use tools. Machines and algorithms build everything from skyscrapers to cell phones. Are they responsible for what they build? While animals and robots are normally considered in distinct philosophical fields, Vogel's rejection of the natural-artificial split prompts us to question the distinction between natural and artificial agents. I argue, under consistent reasons, that neither animals nor robots are morally responsible for what they do. When machines act in morally consequential ways, then, we cannot blame the robot. However, we usually think to blame those who built the robot. I present a theory of how a builder may be responsible for what they build. Then, I argue that there are cases where neither the robot nor the engineer can be blamed for the robot's actions. Drawing on Vogel, Karl Marx, and Martin Heidegger, I explore moral and environmental responsibility through meditations on animals and machines.

For Diane, Toni, Ana, and my mom.
Four women who have shown me unending guidance, care, and, above all, patience.

Contents

1. Building like a Mall 1

1.1 Thinking like a Mall 2

1.2 Contradiction 3

1.3 Heidegger 10

1.4 Marx 13

1.5 Ethical & Political Implications 16

2. Moral Agency in Animals & Machines 26

2.1 Persons & Reasons for Moral Responsibility 28

2.2 Moral Animals 30

2.3 Responsible Robots 38

2.4 Animals & Robots 44

3. Building & Moral Responsibility 56

3.1 Vogel on Building 57

3.2 Building as Cultivation 63

3.3 What is building? 64

3.4 Building & moral responsibility 66

3.5 When are we not responsible for what we build? 71

3.6 Moral Responsibility for Built Agents 74

Works Cited 83

Chapter One: Building like a Mall

Nature has ended. Nothing today is untouched by human hands. Global warming and acid rain blanket the earth's surface. Nuclear waste and fossil fuel extraction alter the earth down to its geophysical foundations. After Bill McKibben's (1989) monumental work *The End of Nature*, environmental philosophy is unable to look to 'nature' as a guide. Furthermore, most environmental scientists and philosophers working today foresee a rapid, catastrophic change in the environment as we know it. McKibben writes that "our reassuring sense of a timeless future, which is drawn from that apparently bottomless well of the past, is a delusion" (1989, p.2). There is an urgent need to adopt a new environmental philosophy that can consistently and powerfully grapple with the environment after the end of nature.

In *Thinking Like a Mall*, Steven Vogel reinterprets the 'environment' as 'that which environs us'. This shifts 'the environment', which is usually tied at the hip with 'nature,' to an environment which is built through human practices. Consequently, environmental ethics cannot rely on 'nature' as a moral standard. Instead, Vogel argues that humans are responsible for the environment because we built it. In a few minor passages, however, Vogel acknowledges that the environment is built through the practices of other organisms and through physical processes. To prevent anthropocentrism, he claims that we build the environment just as any other thing does. In this paper, I argue that this contradicts his central thesis of environmental responsibility. In short, if building implies responsibility and everything builds the environment, then everything must be just as responsible for the environment as humans are. In order to get out of this contradiction, I argue, Vogel's postnaturalist environmental philosophy needs to make a stronger division between the ways that humans and others build.

I will begin with Vogel's notion of *thinking like a mall* and how it relates to environmental responsibility. Then, I will discuss how everything builds the environment and illuminate a contradiction in his work. To remedy this contradiction, building must be considered as ontologically different between humans and others. To this end, I will discuss some of Vogel's understanding of building and suggest two possible clarifications of his philosophy that avoid this contradiction: 1) from Heidegger on building and

thinking, and 2) from Marx on production and species-being. These two proposals salvage some essential difference between humans and others, while rejecting the ontological distinction between nature and artifact. Each is compatible with Vogel's central thesis of anthropocentric environmental responsibility in the built environment.[1] Furthermore, Marx's solution has important implications on Vogel's understanding of environmental responsibility and politics. I will conclude by discussing some political implications of my solution and how they relate to the Anthropocene.

Section One: Thinking Like a Mall

'Thinking like a mall' pays homage to a passage in *A Sand County Almanac* where Aldo Leopold is hunting on a mountain and shoots a mother wolf in front of her cubs (Leopold 1949). His troupe approaches the injured wolf only "'to watch a fierce green fire dying in her eyes" (Vogel 2015, p.129). Leopold "saw in that fire something 'known only to her and to the mountain'" (p.129). For Leopold, 'thinking like a mountain' is to think beyond our limited understanding as humans in order to recognize some deep, mystical knowledge held in 'natural' things. Conversely, Vogel asks: if we can *think like a mountain*, why can we not *think like a mall*? '*Thinking like a mall*' challenges the "dark complexity and depth of the processes of nature" (p.129). It considers the mall, an artifact that is not living, as a subject of environmental philosophy. It takes environmental as political. It makes no ontological distinction between the natural and the artificial.

To *think like a mall* is to consider how the environment is more like a mall than a mountain: we build the environment. And because we built it, Vogel writes,

> *[w]e are responsible for the environment*: this [is] the central conclusion of… postnaturalism in environmental philosophy. The environment itself is an artifact that we make through our practices,

[1] There are many other ways to define a moral agent in such a way that precludes animals. Aristotle suggests that animals are not rational nor do they have language, which are necessary for moral agency. Frankfurt (1971) suggests that animals do not have second-order volition. Many of these derive from the analytic tradition. Here, I give definitions of personhood and moral agency from Heidegger and Marx (1) to approach this topic in the Continental tradition and (2) to give an in-depth reading of Vogel's work. Vogel relies heavily on both Heidegger and Marx. As well, others have written on connections between Vogel and Heidegger (see Magdalena Hoły-Łuczaj's forthcoming work, "Revisiting an Ecophilosophical Reading of Heidegger").

> and hence one for which we are responsible and about which we ought to care. If the artifacts that surround us are ugly, if they work poorly, if they generate poisons and other toxic waste, if they make life worse for us and for the other creatures that inhabit the world with us… this is (in part) our doing, and our fault. And so it is also our responsibility to fix (Vogel 2015, p.164-165, emphasis in original).

The above section illustrates Vogel's central thesis of his work: building implies responsibility. When we make artifacts ugly or they harm other creatures, we are responsible for making them better. The environment is an artifact. Humans have made it ugly and it harms other creatures. Thus, we ought to make it better.

However, this isn't exactly what Vogel says. In fact, he says that the state of the environment is our doing, but only *in part*. There are many other things that build the environment. Vogel writes that "the environment comes to be what it is *through* our practices, just as it comes to be what it is through the actions of beavers, honeybees, earthworms, trees, and all other organisms that make up the worlds," as well as non-organisms like "water and soil,... subways, airplanes, and incinerators" (2015, p.66, emphasis in original). He makes a few short claims like this throughout his work --saying that we cannot make any meaningful distinction between how humans build the environment and how everything else builds. This is a secondary claim, but it stands out. Vogel holds that humans are no different from any other organism or thing when it comes to building the environment. But, if humans and other organisms all build the environment, then why is it that every other organism or thing is not just as responsible as humans are?

Section Two: Contradiction

To say that everything builds undermines Vogel's central thesis of human responsibility. Vogel's entire argument rests on his ontology of building. To say that the environment *comes to be what it is* through practices of beavers, water, or subways is to say that all other beings build the environment as humans do. Thus, if we accept his central thesis --that building implies responsibility,-- it must follow that all these others organisms and things are also responsible for the environment. But, it seems absurd to say that beavers or water are responsible for global warming in the same way that humans are. There are so

many other things that build the environment that it would be impossible to pinpoint what humans are uniquely responsible for and what everything else is responsible for. His central thesis becomes muddled if we allow that everything builds the environment with equal responsibility.

To make this contradiction and its implications clear, we need to distinguish between moral and causal responsibility. Causal responsibility is simply assigned to the effect that made brings a cause into existence. For example, suppose a domino falls on its own and hits another domino, which hits another and another down a long string of dominoes. When the last domino falls, this event was brought into existence by every domino in the string; though, it was brought into existence primarily by the first domino. Thus, the first domino is causally responsible for the fall of the last domino; though every other domino along the way is also causally responsible as well, to a lesser degree.

Now suppose that, when the last domino falls, it hits a baby and wakes it up. The baby starts crying. When the parent comes into the room, they can see that the first domino fell on its own and is responsible for waking the baby, but they do not scold the domino. The domino is not morally responsible for the events that it causes; only causally responsible. Alternatively, if the baby's older brother had pushed the first domino over, then the older brother would be not only causally, but morally responsible for his actions.

The contradiction that I laid out above applies to moral responsibility. When Vogel says that "[w]e [humans] are responsible for the environment," he means that we are *morally responsible* for it: this is the central conclusion of postnaturalist environmental philosophy (2015, p.164). Thus, I argue that Vogel's central thesis becomes muddled if we allow that non-humans things are also morally responsible for the environment. To clear Vogel's contradiction, we must claim that when the environment comes to be what it is through the practices of non-human things, like rivers, subways, or bees, those things are only causally responsible for the environment they build. They are not morally responsible.

This is a common viewpoint; not one that Vogel would want to disagree with. However, saying that only humans are morally responsible for their actions necessitates an ontological distinction between humans and natural organisms. Vogel argues against ontological distinctions between things that are built

by humans and things that are not, i.e. artificial and natural. When he writes that "the environment comes to be what it is *through* our practices, just as it comes to be what it is through the actions of [others]," it is not entirely clear whether he makes an ontological distinction or not. In some sense, one could read the phrase "*just* as it comes to be what it is" as simply recognizing that other things also build the environment; yet, they build in fundamentally different ways so as to make only humans responsible for their actions (Vogel 2015, p.66, emphasis mine).

However, it is also plausible that Vogel wants to reject any ontological distinction between humans and natural organisms grounded in building. In other passages, Vogel writes that claiming that humans are responsible for the environment sounds a lot like claiming that we are "masters of the world" (2015, p.164). In order to prevent this "sort of hubris," Vogel emphasizes humility as an environmental virtue (p.164). Similarly, Vogel attempts to rid of any anthropocentric hubris that comes with making ontological distinctions between humans and natural things. Thus, the "*just as*" in the passage "*just as* it comes to be what it is" can also reasonably be read as a rejection of the ontological distinction between humans and natural organisms, like animals, based on how we build the environment (p.66, emphasis mine). In this case, there are no grounds that allow Vogel to assign moral responsibility to humans and not natural organisms.

In other passages, he makes clear that even if "the environment comes to be what it is *through* our practices, just as it comes to be" through the actions of others, this is not to mean that everything builds equally in any shallow sense (2015, p.66, emphasis in original). Beavers do not use backhoes. Subways cannot build houses. Honeybees cannot craft atomic bombs.

Rather, Vogel writes that

> we can distinguish between things that humans make and things they don't --they make toasters and shopping malls, they don't make rocks or mountains-- and we can even distinguish among various 'degrees' of human-madeness --there's much more human work involved in the construction of a skyscraper than in the building of a hut in the Black Forest... But none of these distinctions has any *ontological* significance: there are doubtless occasions where it might be useful to draw them so as to indicate the relative role played by different organisms in an item's genesis, in the same way, for example, that one might want to distinguish soil that has been well fertilized

> and aerated by earthworms from soil that has not, but there is nothing ontologically fundamental implied by these distinctions (Vogel 2015, p.169, emphasis in original).

Here Vogel says that, given an object, we can distinguish between how much humans make it and how much other organisms make it. In this passage, he is not explicitly writing about moral responsibility. However, he does say that "there are doubtless occasions where it might be useful to draw [distinctions] so as to indicate the relative role played by different organisms in an item's genesis" (p.169). One may read this and respond that assigning anthropocentric moral responsibility for the environment is an occasion where it might be useful to draw a distinction so as to indicate the relative responsibility played by different organisms. In this way, we can determine how responsible each organism is for making an object. A worm is causally responsible for aerating a pile of soil, but a human is not. For the case of the environment, if we can determine how much humans make it and how much other things make it, we can determine how much each thing is responsible for the environment. However, this distinction is not *ontologically significant*. Thus, it seems that we can read Vogel's philosophy as ridding of an ontological distinction between humans and nature, while allowing for a moral distinction to be drawn. It seems that Vogel's philosophy does not run into the contradiction outlined above.

However, this reading is flawed in two ways: 1) assigning moral responsibility only to humans requires some ontological significance between the distinction between how humans and non-humans organisms build; and 2) he falsely reduces the task of assigning responsibility for the environment to the simplistic task of assigning responsibility for a single object.

First, even if we can assign relative responsibility for building different things, this still does not reveal on what grounds Vogel makes the claim that only humans are morally responsible. We can tell that worms aerate a handful of soil. They are causally responsible for it. Yet, if that soil is built badly, the worm is not morally responsible for improving it. We can also tell that humans build a mall. They are causally responsible for it. However, if that mall is built badly, the people are morally responsible for improving it. This is the distinction that Vogel makes. It is a moral distinction that is grounded in an ontological distinction between humans and others.

Yet, he argues that there is no *ontological significance* to distinctions between things that are human-made and things that are not. *Ontological significance* refers to something like the deep, mystical knowledge in Leopold's mountain. It is something special about 'nature' that exists independently of humans. For example, many environmental thinkers treat ecologically restored habitats as second-class to 'natural' habitats that are otherwise identical but were not regrown by humans.[2] They are lesser because they do not have the property of being 'natural'. This is simply a point about the things that are built. It is not a point about the builders themselves.

Vogel's philosophy needs to make an explicit some ontological distinction between humans and others so as to assign moral responsibility solely to humans, in order to avoid the contradiction outlined above. He rejects the ontological significance of distinctions between natural and artificial *things*. Thus, he must make an ontological distinction grounded in how *builders* --humans and other organisms-- are different. This requires an explanation of how humans and other organisms build differently.

Second, it requires greater distinction to assign human responsibility for the environment than it does to assign human responsibility for an individual object. We can say that humans build toasters or even shopping malls. We can also say that they don't make rocks or mountains. But these are tangible, everyday things. The environment is not tangible and cannot be understood as an individual object like a toaster or a mall.

Timothy Morton would describe the environment as a *hyperobject*: a thing that is "massively distributed in time and space relative to humans" and cannot manifest itself locally (p.1). (2013). For example, Morton also describes the Solar System, all of the plutonium on earth, and the Florida Everglades as hyperobjects. Because hyperobjects do not manifest themselves locally, they "exhibit their effects *interobjectively*; that is, they can only be detected in a space that consists of interrelationships between [individual] objects" (p.1). For example, one can feel heat and radiation from holding a single stick of plutonium. However, only in Nagasaki or Chernobyl can we detect the effects of an enormous

[2] See Robert Elliot's "Faking Nature" (1982) or Eric Katz' "The Big Lie" (1992).

collection of plutonium over a long period of time. Furthermore, the effects of the totality of plutonium can only be detected through the largest possible generalizations from effects of single sticks or enormous collections. The environment is similar in that it can be observed only by first making general claims about weather patterns over time (climate), then about how climate changes over time, and finally about how local climate changes add up globally. Thus, determining how much we built the environment must be determined through large, scientific, statistically-mediated means. We can take millions of local temperatures and samples of sea ice over thousands of years, measure all the added carbon emissions through cars and agriculture and deforestation, then make a scientific theory to explain the raising temperatures and shrinking sea ice via greenhouse gases. This is very different from reading "Made in China" on the bottom of a toaster.

The enormity and vagueness of assigning causal responsibility to the environment requires greater demarcation between humans and others than assigning causal responsibility to an individual object does. The enormity and vagueness of how the environment is affected by the practices of anything requires makes assigning causal responsibility for the environment more difficult than assigning causal responsibility for an individual object. Since even causal responsibility is more unclear, a philosophy that claims that humans are solely responsible for the environment cannot waver on its distinctions between humans and other organisms in assigning moral responsibility. If Vogel allows that animals build similarly to humans, then it would seem that these animals are similarly morally responsible for the things that they build. On an environmental scale, even the smallest bit of moral responsibility for things is exacerbated by orders of magnitudes, due to the sheer amount of things that are built or influenced by animals. If Vogel allows for similarity between how humans and others build, this would lead to a muddling of moral responsibility on an environmental scale. Thus, he needs to make a clear ontological distinction between humans and natural organisms grounded in building. He needs to say exactly why humans build in a unique manner in order to claim that only humans are morally responsible for the environment.

To be clear, Vogel does make distinctions between humans and nature based on language. He writes "[n]ature does not… actually use language, because to use language is to converse" and "nature does not talk with us, it talks at us" (p.185). Traditionally in environmental philosophy, when nature is described as having language, it is never capable of listening; rather

> we respond to it like silent subjects listening to the commands of a monarch, not like participants in a dialogue who develop mutual understanding and respect through repeatedly and alternately taking up the positions of speaker and listener (p.185).

Vogel argues that nature is incapable of engaging in conversation and, thus, does not have language. Furthermore, since justification can only arise in conversation, nature cannot make justifiable claims: only tautologies. Ethical claims are necessarily justifiable and arise in dialogue. Thus, nature does not have an ethics. Only humans, who have language, can have ethics.

At first glance, this seems like a way for Vogel to say that only humans are morally responsible. One could argue that, since only humans can engage in conversation that leads to moral assertions, only humans can assert who is morally responsible.

Yet, this is lacking. This would only imply that humans can determine who is blameworthy, not that only humans are blameworthy. Someone can be held morally responsible even if they cannot engage in conversation. For example, consider two opposing soldiers that do not speak the same language; say, a Vietnamese soldier and an American. The American shoots and wounds the Vietnamese soldier. Even if they cannot engage in conversation, the Vietnamese soldier still holds the American morally responsible for his wound.

Furthermore, simply having ethics is not sufficient for assigning moral responsibility. For example, suppose that the American and the Vietnamese soldiers have contradictory ethics based in their respective languages. The Vietnamese soldiers still blames the American for shooting him. Yet, it is not because they operate under the same ethics. They both have the capacity for ethics and language. However, the use of language or ethics is not sufficient for assigning moral responsibility.

Thus, any distinction between man and nature that implies solely human moral responsibility must do so through deeper means than surface-level language use, like engaging in conversation. To say humans are responsible for the environment because we built it requires demarcating humans from nature *through building*.

Section Three: Heidegger

Heidegger offers a solution to Vogel's contradiction by distinguishing building between humans and animals through thinking and language. Heidegger grounds the fact that only humans use language in a deeper ontological difference between humans and nature. Considering Heidegger's understanding of building in Vogel's philosophy will allow for a way to clarify how Vogel assigns moral responsibility solely to humans. This avoids the contradiction that arises when considering natural organisms as morally responsible in Vogel's philosophy. Furthermore, Heidegger's solution fits well within Vogel's philosophy because Vogel draws heavily from Heidegger and both of their understandings of language-use are very similar.

Heidegger, like Vogel, also draws a distinction between humans and animals through language and dialogue.[3] *Dialogue* is important to both Heidegger and Vogel in how they conceive of language and *logos*, which can be read as synonymous. For Heidegger, "*λόγος* (*logos*) as discourse really means *δηλοῦν* (*deloun*), to make manifest 'what is being talked about' in discourse" ([1953] 2010, p.30). Furthermore, "*λόγος (logos)* lets something be seen" in that engaging in dialogue discloses truth.

Both thinkers follow the classical Aristotelian alterity relation with animality where man is *zoon logon ekhon* and animal is *alogon* (without *logos*).[4] Rather than interpreting *logos* as logic, reason, or rationality, both thinkers interpret it as *speaking* and, in particular, *dialogue*. As well, for Heidegger thinking is "like building a cabinet:" "it is a craft, a 'handicraft'" (Heidegger 1976, p.16). Thus,

[3] I consider animals as a paradigmatic case for natural organisms. If we draw an ontological distinction between humans and animals, then we can draw one for any natural organism.
[4] For Heidegger, animals are poor in a number of different ways: they are "without language, without history, without hands, without dwelling, without space" (Elden, 2006 p.274).

Heidegger connects thinking and building like Vogel does (*thinking like a mall* is a meditation on building). However, Heidegger articulates this connection through *hands*:

> Every motion of the hand in every one of its works carries itself through the element of thinking, every bearing of the hand bears itself in that element. All the work of the hand is rooted in thinking (Heidegger 1976, p.16).

From the above passage, we see how intimately connected the hand and thinking are. Vogel considers building and thinking first, then considers language as a means of placing ethics particularly on humans. Conversely, Heidegger's understanding of building as *handi*-craft ties building, thinking, and language together. It also reveals a way to ontologically demarcate between humans and animals within Vogel's philosophy. On animality, Heidegger writes:

> In the common view, the hand is part of our bodily organism. But the hand's essence can never be determined, or explained, by its being an organ which can grasp. Apes, too, have organs that can grasp, but they do not have hands. The hand is infinitely different from all grasping organs --paws, claws, or fangs-- different by an abyss of essence. Only a being who can speak, that is, think, can have hands and can be handy in achieving works of handicraft (Heidegger 1976, p.16).

Heidegger thinks that apes cannot grasp because they cannot speak and cannot think. This is just like how Vogel thinks that apes (or any natural things, for that matter) cannot have ethics because they cannot speak or have dialogue.[5] The hand is a particularly human thing that differs from animals' grasping organs based on thought and language. They also allow us to build in a unique way, as handi-craft.

If we consider Heidegger's hands in Vogel's philosophy, then we can separate humans and nature in the ways that they build the environment. From here, we can interpret hands as carrying a particular responsibility, as opposed to the lack of responsibility that comes with the simple grasping organs of animals or the complete lack of hands for any other natural thing. Heidegger does not explicitly connect hands and responsibility in this way. However, we can argue that because we have hands, we have a

[5] The necessity of language for morality in animals is also disputed. See Bekoff and Pierce (2009), Shapiro (2006), or DeGrazia (1996). I will discuss this more in full in the next chapter.

unique responsibility for the things that we build. I will refrain from giving too close of an analysis of Heidegger's hands and why we can assign responsibility through them, for it lies beyond the scope of this paper. On the surface, however, we are responsible for our handi-craft because we think and are intentional about how we build. Animals do not think about the things they build, thus they are not morally responsible for them. Thus, if we accept that humans have hands and animals do not, then we can say that when humans build the environment, they are responsible for it; when animals (and all other natural things) build the environment, they are not.

Above, I outlined a couple of ways that Heidegger and Vogel seem to agree on their understandings of language, thinking, and building. And thus, hands seem to be a natural fit into Vogel's philosophy. However, upon further analysis, Heidegger's understanding of thinking and building may be too contradictory for Vogel's philosophy to accommodate for 'handi-craft'.

In *In the Swarm*, Byung-Chul Han notes that "Heidegger removes the hand from the sphere of action and situates it in relation to thought. Its essence is not *ethos* but *logos*" (Han 2017, p.38, emphasis in original). Because "[a]ll the work of the hand is rooted in thinking" (Heidegger 1976, p.16), the hand acts, but it does not act "in terms of *vita activa*" (Han 2017, p.37, emphasis in original). In other words, "Heidegger's hand thinks *instead* of acting" (Han 2017, p.37, emphasis mine).

This is the opposite of Vogel. For Vogel, *thinking* (as in '*thinking like a mall*') derives from *building*. For Vogel, "[p]ractices are our way of coming to know the world --they are… our way of being-in-the-world" (Vogel 2015, p.56). Our practices are always changing and building the world around us. Furthermore, the ways people think (and in particular how they think about nature) "in hunter-gatherer or agricultural or ancient urban or feudal or industrial or postindustrial societies… are expressions of the fact that… [these] societies *engage in different kinds of practices*" (p.57). For Vogel, thinking forms ideal intentions. It does not build material things. Rather, the complete opposite: our practices inform our ideals.

Where "Heidegger's hand *thinks* instead of acting," Vogel's hand *acts* instead of thinking (Han, 2017, p.37). Vogel's hand would have to manifest itself as a manual operation to build a material thing,

like a crew of workers literally constructing a mall. However, Heidegger's hand does not "manifest itself as a manual operation [*Handlung*] but… as handwriting" (Han 2017, p.37). Vogel's hand *acts* in terms of *vita activa* instead of writing. This may be grounds enough to reject Heidegger's hands as a possibility to adopt into Vogel's philosophy. Thus, it seems that a more suitable alternative to provide a solution to Vogel's contradiction of building and responsibility must conform to Vogel's materialism. We turn to Marx to provide such a solution.

Section Four: Marx

Vogel draws from Marx's labor by taking building as a social, material practice. We build the environment like a crew of workers builds a mall. We emancipate ourselves from alienation when we understand that the environment is built in this way. Vogel's thesis of material construction arises from the notion that we are *alienated* from the environment (i.e. 'nature'). Usually, this means that modernity has lost a connection with nature; thus, we "ignore the impact of our actions on it" (Vogel 2015, p.66). Environmental ethics, under this view, ought to restore our connection with nature by making sure that more people go camping in pristine forests and sparkling rivers to realize that these 'natural' things are more than commodities. Vogel returns *alienation* to Marx in order to critique the notion of nature and derive an environmental ethics and a politics without it.

Marx (drawing on Hegel) writes that the worker is *alienated* from their labor, when "the object which labor produces, its product, stands opposed to it as *something alien*, as a *power independent* of the producer;" as a physical manifestation of the power that the property owner has over the worker (Marx [1844] 1975, p.324, emphasis in original). "The worker is related to the *product of his labor* as to an *alien* object" (p.324, emphasis in original). For example, if Jane spends an afternoon sanding and squaring off the corner cuts of a cabinet, this cabinet is bound only to be separated from her by the owner of the of the cabinet-making shop. The cabinet appears to her as "*loss of realization*," as "*loss of the object and bondage to it*; appropriation [by the owner] as *estrangement*, as *alienation*" (p.324, emphasis in original).

To be alienated from the environment, then, is not to be separated from 'nature': to live on pavement instead of prairie grass. We are alienated from the environment when it appears to us as a loss of realization. We do not realize our own labors produce "the environment in which we live, the homes and offices within which we work," and even "the very clothing that we wear and the food we eat" (Vogel 2015, p.86). Those of us that produce might be able to recognize some of the individual things we produce as our own. Collectively, however, we do not recognize our actions in the environment as a whole. Each time we drive a car and emit carbon dioxide into the atmosphere. Each time someone eats fried chicken, the entire supply chain from egg to farm to slaughter to table, all the workers and emissions associated, affect the environment. Rather than recognizing that our actions deeply affect how the environment is built, we only see our private actions: driving to work in the morning, eating chicken at lunch, etc.

Thus, "[i]t is precisely in this failure of humans to 'see themselves in the world they have created' that their alienation consists" (Vogel 2015, p.72). To consider the environment as built is to realize ourselves in nature and emancipate ourselves from alienation. Nature is apart of human practices and, conversely, humans and practices are apart of nature. To recognize the full extent of our practices and our labors is to understand that nature is an obsolete term. This is what is encapsulated in considering the environment as built, i.e. *to think like a mall*. This is how to emancipate oneself from alienation from the environment.

Just as alienated labor forms Marx's understanding of labor, alienation from the environment forms Vogel's understanding of building:

> To see humans as part of the world, as entwined with the world, would be to see in each object in one's environment a history of human practice, and at the same time to recognize that humans don't think or intend or imagine or perceive or reason or even somehow magically constitute the world but rather *engage in practice within* the world (Vogel 2015, p.94, emphasis in original).

Here, Vogel outlines a way to emancipate ourselves from alienation: through an understanding of building as practice. Specifically, Vogel considers building as a social, material practice; he draws this from Marx's notion of *labor*. Marx's writing on alienation and labor directly forms Vogel's

understanding of building. Thus, Marx gives us a good second solution to ontologically demarcate humans and animals in Vogel's philosophy.

In precisely the passage from the *1844 Manuscripts* that Vogel draws from to understand alienation, he omits Marx's discussion on human nature and animality. Marx writes that

> animals also produce. They build themselves nests and dwellings, like the bee, the beaver, the ant, etc. But they produce only their immediate need or those of their young; they produce one-sidedly, whilst man produces universally; they produces only when immediate physical need compels them to do so, while man produces even when he is free from physical need and truly produces in freedom from such need; they produce only themselves, while man reproduces the whole of nature. (1975, p.329).

Furthermore, Marx declares that, through production "man really proves himself to be a *species-being*" (1975, p.329, emphasis in original). Marx draws the line between humans and animals based on how we each produce based on *need*, *universality*, and *species-being*. Animals, in opposition, produce only for *immediate need*; they produce *one-sidedly*, for themself or their young. For example, a hyena tears flesh off a carcass to feed itself. It does not act beyond its immediate need. It feeds itself only; not its peers; especially not any other animal or thing beyond its own species. Thus, it only produces one-sidedly. For Marx, this is essentially how all animals produce.

Humans, on the other hand, have the capacity to produce beyond their immediate need in accord with their species-being. Marx writes that "free conscious activity constitutes the species-character of man" (p.328). Whereas the animal "is immediately one with its life activity" and, like the hyena, only produces to sustain its own life, man "makes his life activity itself an object of his will and consciousness" (p.328). Thus, humans are conscious of our own actions in a way that animals are not. This makes us able to produce, even when we are *free* from immediate physical need.

Furthermore, producing free from need is not simply a way of producing based on your private whims. Humans are conscious of their own species-being when they produce in accord with both their individual humanity and with a larger collective body within which they work. In *On the Jewish Question*, Marx writes that man will "become a *species-being*" when "the real, individual man has absorbed into himself the abstract citizen" (1972, p.46). When you produce like an *abstract citizen*, you

produce for the good of society and of humanity. Beyond that even, Marx writes that "man reproduces the whole of nature" in that man can consider all of nature when he produces. Thus, humans are capable of producing when they are free from immediate need in such a way that they are conscious of the collective social, or even biotic, context of their labors. This sets human practices apart from those of animals.

If we incorporate building as production into Vogel's philosophy along with Marx's notion of species-being, we can argue that humans are uniquely capable of being morally responsible for the things they build. Even after discussing species-being above, it may not be entirely clear what Marx means by this piece of jargon. I read Marx simply as saying that humans are capable of being conscious of their practices both at an individual level and with regards to a larger social context, e.g. their neighborhood, their community, their species (humans), other species, or even the whole of earth. In fact, the individual and the social are not separate for Marx: he writes that "[w]hen man confronts himself, he also confronts *other* men" (1975, p.330).[6] Because humans have this capacity to understand the social implications of what they build, they have moral obligations to others that are affected and they may be morally responsible for what they build. Animals, as well as all other natural organisms, are not capable of understanding their practices in a social context. Thus, they are not morally responsible for what they build. Solely humans are morally responsible for what they build.

This solution is preferable to Heidegger's because it fits best with Vogel's understanding of building as a social, material practice. Furthermore, adopting Marx's understanding of building as producing in accordance with species-being into Vogel's philosophy has nice implications for an ethics and a politics that align very well with those of Vogel.

Section Five: Ethical and Political Implications

There seems to be a bit of a hitch with adopting Marx's solution forthright. Marx considers production in accordance with species-being (read: building in a social context) an ideal situation. It is

[6] This is rather like in *Sources of Normativity* where Christine Korsgaard (1996) argues that one has obligations to others and to humanity simply based on one's capacity to see oneself as human (p.143).

more of an eschatological ideal than an actual reality. Under our current economic organization, our labor is alienated. Workers do not offer their labor in order to produce things that benefit themselves and their society. They do so because they need to get a wage. Their labor power is converted into an object --a product-- and stripped away from them by their boss or sold on the market in order to get money in return. The objects that workers produce confront them as alien powers: as reminders of their wage-labor relationship and the loss of realization of their labor power.

For most, this means that you get up and produce only so that you can get enough money to pay for rent and food: basic means of staying alive. Whereas the emancipated man is a conscious being in that "his own life is an object for him, only because he is a species-being" and that "his activity [is] free activity," alienated labor "reverses the relationship so that man, just because he is a conscious being, makes his life activity, his *being* [*Wesen*], a mere means for his *existence*" (1975, p.328). Thus, producing for wages drives workers to produce for immediate need.

This makes it so that workers do not realize the social nature of the things that they produce. Marx writes that a consequence of "man's estrangement from the product of his labour, his life activity, his species-being, is *estrangement of man from man*" (p.330). Thus, not only does alienation cause a worker to produce only for themselves, but they lose the capacity to realize that their labor is connected to a larger society as a whole.

In all, alienated labor reduces the human capacity for production to its animal base. Marx writes: "In general, the proposition that man is estranged from his species-being means that each man is estranged from the others and that all are estranged from man's essence" (p.330). Alienated labor "transforms his advantage over animals into the disadvantage that his inorganic body, nature, is taken from him" (p.329). When humans produce for a wage, they produce only for "*greed,* and the *war of the avaricious – competition*" (p.323). This is no better than the hyena. Alienation reduces human production into private, selfish affairs that only sustain life for the worker. As a result, the worker "feels that he is acting *freely* only in his animal functions --eating, drinking and procreating, or at most in his dwelling and adornment-- while in his human functions *he is nothing more than an animal*" (p.327, emphasis mine).

This seems like a hitch in my formulation of Marx's definition of moral agency, i.e. an agent capable of producing in accord with species-being, because it seems like it would excuse all humans that produce under alienation of moral responsibility for what they build.[7] Let me sketch this out here. Blame is assigned only to agents who produce within a social context. Animals do not have moral responsibility because they do not produce in accord with species-being. Humans are reduced to the level of animals when they produce under alienation. Thus, it seems as though humans would be as morally responsible for the things that they create under alienation as animals are. That is to say, humans would not be morally responsible at all for their environmental practices (like animals).

This is partially true. Humans build the environment at a collective scale. Our societies, their economic organization, their geographical organization, and their cultural standards lead us to act in ways that create the environment. We don't decide, on an individual level, these larger ways of organizing socially. This is a point that Vogel explicates well. He writes:

> Each morning I make an implicit decision about whether to drive to work or not: driving generates a certain amount of carbon dioxide and therefore helps to increase the future temperature of the planet. Not driving would cause me to lose my job (Vogel 2015, p.203).

This problem has to do with the "private character of the decision with which [he is] faced" (p.202). "Operating within a market economy, [people] have to act as private individuals whose acts are independent of the actions of others" (p.202). Our actions are made to seem private when, in fact, they add up to build the environment. "Global warming is a social product" (p.202). Yet, it is an alien power that arises when we do not recognize the sociality of our practices. This is "the structure that Marx described under the name of alienation" (p.202).

Because our practices are made to be private by the social and economic conditions we live in, we are ignorant of the particular consequences of our environmental actions. Only the best climatologists understand the impact of the carbon they emit when they drive to work; and, even then, they can only

[7] This does not mean that people that produce under alienation are not culpable for any of their actions: if you work for a wage and you stab someone out of cold blood on the street, you should still be blamed for murder. I am mainly thinking of things you build when you produce for a wage under a boss.

understand the impact of their emissions indirectly, within a collective of toxic commuters. These actions are not intentional. Nobody wants global warming. They simply accept it because they want food and shelter and energy and entertainment and the ways we get those things currently comes with a steep environmental price. Finally, these actions are not entirely voluntary.[8] Vogel does not have a choice whether or not to drive to work. He alone cannot rebuild the entire highway system, create more public buses, or move his work closer to his neighborhood.

To be blamed for one's actions, a moral agent must (1) not be ignorant of the particular consequences of their actions, (2) they have to act somewhat intentionally (in a non-negligent manner), and (3) they have to act voluntarily.[9] However, as I argued above, individuals that build the environment through their private, alienated practices do so ignorantly, unintentionally, and involuntarily. Thus, one cannot be blamed for their individual environmental practices. For example, we ought not to blame Vogel for the environmental harms caused by driving to work. Walter Sinnott-Armstrong (2005) would argue the same: that an individual has "no moral obligations not to waste gas" when they drive to work in the morning (p.312). Sinnott-Armstrong argues this for different reasons. He thinks the consequences of our individual actions are negligible. While I don't entirely agree that the environmental consequences are null, I do think that Sinnott-Armstrong captures a powerful point about the very small, practically imperceptible impacts that individuals have within our environment.

Though our individual actions are, to an extent, involuntary --like Vogel being forced to drive to work in the morning,-- Sinnott-Armstrong argues that "*governments* still have moral obligations to fight global warming, because they can make a difference" (p.312, emphasis mine). Sinnott-Armstrong is arguing not that we ought to blame governments; but, rather that they have a forward-looking responsibility to make the environment better, because they have the power to effect this change. This

[8] This point may be contentious with regards to Aristotle's definition of voluntary action. Aristotle only considers actions under the most extreme forms of duress to be involuntary.
[9] I will discuss blameworthiness, ignorance, intentionality, and voluntary action more in Chapter Three.

does not entirely capture the sense of moral responsibility that I have been discussing.[10] However, it does capture the sense in which moral obligations should be assigned on a larger social level.

While our individual actions may seem inconsequential, our collective actions obviously make a difference. We collectively build the environment. However, we do so as an aggregate of our individual actions. Once we recognize the power that we hold as a society --Sinnott-Armstrong would say, that the government holds,-- we have the capacity to collectivize and lead the direction of our society towards more environmentally-sound practices. This is Vogel's notion of "discursive democracy" (p.236).

Even if we cannot assign individual moral responsibility for our environmental practices, I think we can understand environmental responsibility as a kind of indirect, shared responsibility derived from our proximity to the collective. Based on my definition of moral agency from Marx, we are morally responsible if we can understand the impact of our actions on a larger social level beyond simply ourselves and our children. We may not be able to fully understand how our individual actions directly affect the environment, but we can understand how they do when we consider ourselves as part of a society that may change by changing how our society is organized. Thus, I think that those who may understand and are more able to affect how our society is organized ought to be more morally responsible for their environmental actions.

Vogel recognizes this as a democratic process. It seems as though all citizens in an environmentally-harmful democracy may be morally responsible for their society's environmental actions. Yet, not all are equal in a democracy. A factory owner chooses to build a plant on a river because it's cheaper to produce their goods. Individual consumers that buy the factory owner's products are implicitly supporting the factory and its harmful environmental actions. The workers, the consumers, and the factory owner are morally responsible in that they are all citizens in the society that produces a factory that pollutes. They all have the means to realize how the structure of their society leads to harmful environmental actions. They all have the means to collectively organize and lead their society in a

[10] See Jessica Nihlén Fahlquist's 2017 work or Chapter Three for a larger discussion of the distinction between forward- and backward-looking responsibility.

different direction. Thus, they are all morally responsible for the collective actions of their society in that they have the ability to organize and create new policies or social rules that lead to better environmental actions. However, this is not to say that they have equal responsibility. The factory owner, who chooses where the factory is built, has more of a say in controlling the means of production than a factory worker does. On Marx's formulation of responsibility, we may say that the factory owner understands the larger social impact of their actions --since they can see the big picture of how the factory will affect its customers, workers, and the land that it pollutes-- better than any individual worker can. One's moral responsibility for the environment is lessened if they have a lessened say in society. Marginalized people or people in poverty, who do not have a voice in the decisions of their workplace or their government, are much less responsible for the society's environmental actions than privileged, rich citizens with a lot of political clout.

I think that one may be excused for their individual environmental actions, for one, if they are ignorant of consequences. Driving your car and emitting greenhouse gases is like going to the polls, adding your small vote to the pool of millions of other voters. When we vote little individual effect and it's difficult to see exactly where our vote goes, since it is lost in a sea of statistics after we count it. This is the same with environmental practices. However, when we vote for a candidate we approve of, we are still slightly swaying the collective choice of a representative towards that candidate. We can see that our vote is counted towards the right candidate on a larger collective scale. One may be ignorant of the direct consequences and still be able to be blamed for voting for the wrong candidate, knowing fully well that their vote counted against the right one. Similarly, it is reasonable to expect that people at least have a general idea of what global warming or polluting is and can understand their actions, however small and private, as contributing to them. Understanding our actions within a collective is a means to show that we are still responsible for our actions.

This is exactly what Vogel has in mind when he writes of social construction. To say that something is socially constructed usually implies that it is socially or culturally contingent. It is arbitrary

and can change. The logic goes: *since we have socially constructed X, then we can socially construct X into Y*. For example, if race is socially constructed, then our social ideals create hierarchies of power that marginalize certain racial groups; thus we, as a society, can change these divisions into something more equitable. Vogel, however, takes social construction as a physical process: "[h]uman beings effect a change in the world when they transform it through their physical actions, which is to say through their socially organized practices" (2015, p.42). Thus, "to say that X is 'socially constructed' would mean that socially organized human beings have built X, through process of labor" (p.56).

By realizing that the environment is socially constructed and that it was constructed poorly (to create things like global warming or acid rain and massive species extinction) we are able to construct it anew. However, conceiving of construction as *practice* rather than *theory* is to see humans in the environment, but also to see that we are not masters of it: "[w]e do not 'think' of a world and then magically bring it into existence; the world is perfectly real and material, just as real and material as our activities are" (Vogel 2015, p.167). Constructing the environment better is not an easy process. It requires the environmental virtue of humility.

Furthermore, realizing that the environment is socially constructed is a means to emancipate ourselves from alienation. In *On the Jewish Question,* Marx writes that man is fully emancipated when "he has recognized and organized his own powers as *social* powers so that he no longer separates this social power from himself as *political* power" (1972, p.46). For one to recognize their powers as social powers is to recognize the collective ways that we build the environment. It is to recognize that you and everyone else like you has to drive to work in the morning and, over time, that leads to a sizeable amount of pollution and carbon emissions. Only when we are unalienated can we build in a way that is in accord with species-being. Only when we understand our collective environmental impacts can we change them for the better.

One promising way to do this is presented via *the Anthropocene*. 'Anthropocene' is a geological term for "the current epoch in which humans and our societies have become a global geophysical force"

(Steffen et al., 2007 p.614).[11] The Holocene was the last geological epoch which encompasses the "past ten to twelve thousand years as agreed upon by the International Geological Congress in Bologna in 1885" (p.615). In order to understand and acknowledge the ways that humans have altered and continue to alter the planet, a group of leading geologists at the International Geological Congress recommended that the geological community recognize the Anthropocene as fundamentally distinct from the Holocene (Carrington, 2016). Secondary scientific journals and major media outlets spread the news across the internet in order for everyday people to engage with and understand the ways our planet is deeply shaped by our actions, down to its geophysical foundations (see The New York Times' or The Guardian's series on the Anthropocene). It may be considered a *strategic* theory in order to guide public thought on how we ought to care for the environment and attempt to mitigate our detrimental practices.

To justify these claims, epoch demarcation needs to be grounded in the rigorous scientific criteria that the Anthropocene is "functionally and stratigraphically distinct from the Holocene" (Waters et al., 2016 p.1). This requires a solid basis of empirical evidence about changes in the Earth's geological structure and a theoretical understanding of how these facts relate to human interaction. Geologists track seismic activity and the geological makeup of the Earth's crust in order to measure significant changes over the last couple hundred years. These changes in the Earth's geological makeup "began around 1800 with the onset of industrialization, the central feature of which was the enormous expansion in the use of fossil fuels" (Steffen et al, 2007 p.614). Experts also cite nuclear waste and the enormous scale of domesticated chicken bones as major factors. Thus, the Anthropocene weds truth-seeking scientific machinery with the strategic goal of shifts public conceptions on how we relate to the environment.

The Anthropocene is an exercise in understanding the breadth and depth of how human activity has constructed the environment. Humans have shaped the environment so thoroughly that we play a dominant role in the geophysical make-up of the earth's crust. By illuminating our practices in constructing the environment, we become unalienated. Thus, we realize our moral responsibility. The

[11] It should be noted that Vogel's work serves as a critique of the natural-artificial fissure dominant in the geological sciences.

Anthropocene is one of the most promising attempts at understanding the weight of our environmental actions and inspiring collective action to construct the environment better.

Conclusion

Once we recognize the ways in which we have built nature and the ways that the things we have built are natural, we are able to consider how to change the ways we build the environment in order to build it better. In order to do so, however, we first need to realize that the environment is not external to us. When we "mistake the environment for nature, failing to recognize that what surrounds us is not 'natural' at all but rather a world that we have built," we fail to see our responsibility for that environment (Vogel 2015, p.92). However, to have an impact *on* the environment is not like a roof that you stand upon and hammer a nail into, having an impact *on* the roof. It's not an independent natural world that humans impress upon, but do not reside in. The environment is an artifact. To see ourselves in it, actively constructing it is to see ourselves as morally responsible for it. Instead of a roof, Vogel's environment is like the house that we live in. Improving our environment, then, should be like painting the walls or fixing the faucet. We are responsible for our environment as we are responsible for our house.

Furthermore, nothing else is morally responsible for how they build the house. Animals and other natural organisms are simply causally responsible for how they build the environment. Because Vogel's philosophy blurs the lines between natural and artificial, he runs the risk of saying that humans and natural things build in similar ways. This leads to a contradiction. In order to avoid this contradiction, I have laid out two solutions that Vogel can incorporate into his philosophy. The first, from Heidegger, separates humans and animals based on hands. Vogel may take building as handi-craft into order to create a consistent philosophy of moral responsibility. Yet, this solution suffers since Heidegger does not treat building completely materially. The second, more promising solution, comes from Marx. Humans are able to produce free from need with an understanding of their self as human and as a part of a larger

social context, i.e. they are able to produce in accord with species-being. This solution ought to be adopted by Vogel. It also has the implications that environmental responsibility is diminished for individuals that produce under alienation.

Vogel urges that we understand the ways that the environment is built. This comes first by understanding how we build it differently than everything else. My suggestions for demarcating humans and nature based on how they build is largely a technical point; it clears up an inconsistency in Vogel's philosophy. However, it also implies a more consistently humanistic environmental philosophy that fits well with the rest of Vogel's work.

Most importantly, Vogel's work gives us the theoretical tools to understand how the environment is built poorly and how to fix it. This is no small task. It requires humility in the face of material structures: city planning, highway systems, or social hierarchies like race or class. Vogel writes that

> an environmentalism after the end of nature would call not for practices that protect nature (which doesn't exist)... or even for ones that protect the "environment" (because that would mean protecting shopping malls and coal-burning power plants and all the other horrors that environ us), but rather for practices in which the actors *acknowledge and take communal responsibility for* their transformative effects on the world, doing so via the procedures of discursive democracy (Vogel 2015, p.231).

Vogel's philosophy can be summed up very nicely in a quote from a review of *Thinking Like a Mall*: "Instead of environmental ethics urging us to commune with nature, it should thus focus on community organizing and political philosophy" (Jeremy Bendik-Keymer 2016, p.508).

Chapter Two: Moral Agency in Animals and Machines

In the last chapter, I argued for ontological distinctions between humans and animals in order to absolve any non-human organisms of moral responsibility. First, I introduced the notion of the *hand* from Heidegger. The hand is often touted for its incredible dexterity and as a key evolutionary feature that made humans, out of all other species, able to develop advanced tools and technology. Heidegger goes further than physical nimbleness and ties the hand to human's unique capacity to think and converse. Heidegger thinks that no other animal, even our closest cousin, the ape, has hands for they do not have language or thought. Based on this, I argued that humans are solely morally responsible for their actions: that humans are the only ones capable of understanding the full extent of their handi-craft. Thus, out of all organisms who build, only humans are responsible for the building they do.

Second, I introduced Marx's notion of production in accordance with species-being. Species-being is the capacity for consciousness of one's actions on both an individual and a social level; the social can be a consideration within a human society, between species, or about the whole of the environment. Animals do not produce in accord with their species-being. They produce only for the need to survive: this drives them to produce entirely individualistically. Humans, on the other hand, can produce even when they don't need to. They can build things that benefit themselves, their neighbors, their country, the entire human race, or the entire world. They can also build things that harm these groups. Because they are able to be conscious of how their actions play out in a social context, they are able to be held morally responsible for their actions. Animals do not have the capacity to understand their actions beyond themselves or their offspring; thus, they cannot be held morally responsible for their actions.

For anyone that has seen or heard of Koko the gorilla or Kanzi the chimpanzee, Heidegger's characterization of apes seems empirically false. Both of these apes spoke sign language and could express simple thoughts. Moreover, even animals with no linguistic capabilities (at least as far as we can tell) have capabilities that are fundamental to thinking: intentionality, understanding of causality, and

planning.[12] Based on new research into animal behavior, we have strong grounds to doubt Heidegger's claim that only humans have hands.

As well, a strong form of Marx's thesis about animals (that they do not think past themselves and their young) is obviously empirically false. Animals from ants to apes organize in cooperative social communities. Animals have shown to be empathetic and show concern for the interests of others, beyond their offspring. However, there is something to be said about a weaker form of Marx's definition, here humans can have a larger social understanding of their actions and animals can have only a restricted, local understanding of their actions. Even if animals can think beyond themselves and their offspring, most, if not all, cannot consider the interests of other species or of the entire world. This is a more plausible interpretation of Marx's thesis on animality (though it might not be true to Marx's original intentions.)

In either case, Heidegger's or Marx's, the assumptions about animality are beginning to come undone. The last fifty years have seen an explosion of research and attention to the behavior of animals. Our presumptions, in daily life and in philosophy, about animals are coming into question. For example, empirical examples of animal tool use have grown by orders of magnitude since the early 1960's. Jane Goodall's observations of "David Greybeard using a blade of grass to get termites out of a hole" broke the first crack in the facade of the notion of man as *Homo faber*, i.e. the sole species that had "evolved the necessary skills to manufacture and use tools" (Bekoff and Pierce 2009, p.20). More recently, Ashley Shew Heflin's (2011) dissertation and (2017) book on animal construction and tool use argues that many of these recent case studies as indicative of a capacity for technology in certain animals.[13] Rather than treating human and animal constructions as different in kind, Heflin argues that they should be considered different in degree only. Beyond a few cases of simple tool use, robust technological knowledge is now seen in a wide variety of animals, including primates, cetaceans, and birds (especially corvids).

---

[12] Though, what Heidegger means by *thinking* here is not the common notion of *thinking*.
[13] Ashley Shew goes by 'Ashley Shew' rather than 'Ashley Shew Heflin' now. When referring to her (2011) dissertation, I use 'Heflin'. When referring to her (2017) book, I use 'Shew'.

Thus far, I have argued that humans are uniquely morally responsible because they are distinct from animals. Particularly, I argued that we build differently. Yet, Heflin argues that this difference is not a difference in kind, but simply a difference in degree. If this argument proves to be fruitful, my adoption of Heidegger's and Marx's definitions (which are based on building) are incorrect. Then, how are we to assign moral responsibility? If building implies moral responsibility and animals build to a certain degree, then ought we consider that they are also morally responsible to a degree? Or should we simply find better criteria to distinguish humans from animals, in terms of moral responsibility? There are pressing critiques of antiquated presumptions about the differences between humans and animals based on recent case studies on animal behavior. Now, what is to distinguish humans from animals as responsible moral agents?

This chapter will address this concern. I'd like to consider how recent critiques of the human-animal distinction affect animals as responsible moral agents. I will begin by outlining basic notions of personhood and moral responsibility. Then, I will consider morality and moral responsibility in animals. I will introduce a novel argument for moral agency in animals. This is an argument about the principles for moral agency between the fields of animal ethics and robot ethics. I argue that, under the robot ethicists' definitions, certain animals may be moral agents and show that this leads to a reductio ad absurdum. Hence, the standards of moral agency from robot ethics ought to be reconsidered in order to maintain consistency with animal ethics, rather than the other way around. Finally, I will conclude by rejecting moral responsibility in animals and machines.

Section One: Persons and Reasons for Moral Responsibility

Traditionally, only humans have been considered moral. This can be considered as two capacities: moral action and moral responsibility. The latter capacity, moral responsibility, is determined by personhood: only persons can be morally responsible. Moral responsibility is the capacity for an agent to be blamed or praised for their actions. Harry Frankfurt (1971) defines personhood by *second-order volition*. He writes that *first-order desires* are "desires to do or not to do one thing or another" (p.7).

Second-order desires are desires about desires: e.g. if an alcoholic does not want to feel the desire to drink. *Second-order volition* arises out of second-order desires; it is someone's want for a desire to bring them into action. For example, if an alcoholic wants to drink because they want to taste a beer and not because they have a craving for alcohol. Frankfurt writes that having second order volition is "essential to being a person" (p.10). It is also essential to being a responsible moral agent. The criterion for moral responsibility, in this case, is "the capacity for reflective self-evaluation" (p.7). Frankfurt writes that this capacity is absent in animals other than man. Frankfurt's criterion for moral responsibility is *agent-based*, in that it is strictly concerned with the mental state of a moral agent; only beings that can have second-order volition can be moral agents. By *moral agent*, I simply mean a being that can act morally and be morally responsible for one's actions.

John Fischer and Mark Ravizza (1998) give another highly influential account of moral responsibility. Their theory suggests that moral responsibility hinges on a weak form of *reason-responsiveness*, i.e. the capacity for a person to do otherwise based on a reason to do so. A strong form of reason-responsiveness is *regulative control*, i.e. the ability to regulate between a number of alternate possibilities and decide how to act. Fischer and Ravizza argue that a weaker kind of control, *guidance control* "grounds moral responsibility for actions" (p.54). Even in a metaphysically-deterministic system, an agent may have guidance control if the mechanism that drives their actions is reasons-responsive. This means that the mechanism that drives their actions would respond differently if given different reasons. Though this theory of moral responsibility requires a low threshold of consciousness and deliberation in order to respond to moral reasons; it primarily focuses on the *mechanism* involved in acting. Thus, it is a *mechanism-based* account of moral responsibility.

In "Computer systems: Moral entities but not moral agents," Deborah G. Johnson (2006) lists five criteria that "[c]ontemporary action theory typically specifies… for human behavior to be considered action" i.e. "behavior arising from moral agency" (p.198-199). These include:

1) An *agent* with an *internal state* consisting of *desires*, *beliefs*, or other *intentional states*;
2) An *outward event* where the agent moves their body or an extension of their body;

3) The internal state causes the outward event;
4) The outward event has an *outward effect* beyond the agent's body; and
5) The effect is received by a patient (p.198).

In order to determine whether a being is an agent that can be morally responsible for their actions, one would consider whether their behaviors fulfill each criterion on this list. If a being X is not a moral agent and is not morally responsible for their actions, one need only show that their behaviors do not fulfill one or more of the items on this list. In Section 3, I will explain further how Johnson argues against attributing moral agency to computer systems. However, for now I simply adopt Johnson's criteria as standard for attributing moral agency. I prefer Johnson's criteria for moral action because (1) and (3) align with Frankfurt's agent-based criterion and (2) and (4) align with Fischer and Ravizza's mechanism-based criterion for moral agency.

Johnson's criterion (5) requires that an agent's moral action be received by a moral patient. To understand what a moral patient is, I turn to animal ethicists who argue that animals are moral patients. This will lead to a further discussion on the moral status of animals.

Section Two: Moral Animals

There is practically unanimous consensus that moral responsibility is absent in nonhuman animals. However, this does not preclude animals from being considered morally. David DeGrazia (2002) writes that "[i]ncreasingly many people claim that animals have moral status, moral rights, or both" (p.13). To say that an animal has *moral status* is to say that the animal "has moral importance in her own right and not simply in relation to humans" (p.13).

Tom Regan (1983) makes the distinction between moral agents and moral patients in order to argue for the moral status of animals. Regan asserts that moral agents "are individuals who have a variety of sophisticated abilities, including in particular the ability to bring impartial moral principles to bear on the determination of what, all considered, morally ought to be done and, having made this determination, to freely choose or fail to choose to act as morality --however it is conceived of-- requires"

(p.151). Moral *patients*, on the other hand, "lack the prerequisites that would enable them to control their own behavior in ways that would make them morally accountable for what they do" (p.152). They lack "the ability to formulate, let alone bring to bear, moral principles in deliberating about which one among a number of possible acts it would be right or proper to perform" (p.152). Thus, they cannot be held morally responsible for their actions as moral agents can. However, moral patients can be considered "on the receiving end of the right or wrong acts of moral agents" (p.154). This is what Johnson (2006) means in her criterion (5) for action. Regan argues that animals are not moral agents because they do not have the capability to freely choose between moral principles. However, he also holds that animals have inherent value because they are subjects-of-a-life, i.e. that they have some capacity for emotions, intentionality, and self-awareness, among other cognitive capabilities. A generous reading of Regan suggests that an animal need only sentience to be a subject-of-a-life (see Mary Anne Warren's 1987 critique of Regan). Because animals have inherent value, they can be treated as moral patients. This gives animals a moral status.

Regan's grounds for moral responsibility require 1) a sophisticated agent who can enumerate their moral principles, and 2) free action that affords the agent a kind of control. The first criterion is agent-based, similar to Frankfurt's. This criterion is also not posited by Regan alone. Evelyn Pluhar (1995) also argues that the ability to formulate and motivate action through moral principles is the threshold for moral agency. I think this criterion is too strong; it relies on human-like linguistic abilities. There is plenty of evidence of animals --dolphins, whales, great apes-- with the ability to communicate and use language in a way that does not look like human speech and symbols. There is also plenty of evidence of animals behaving in seemingly moral ways without any linguistic ability. I will discuss these in more detail later when considering the works of Shapiro (2006), Bekoff and Pierce (2009), and Rowlands (2012). For now, it suffices to say that a weaker criterion than Regan's can be adopted as an agent-based threshold for moral responsibility.

Regan's second criterion can also be supplanted with a weaker one. This criterion is mechanism-based. Regan would likely endorse a kind of regulative control, which necessitates alternate possibilities

to choose from, as a threshold for moral agency. Instead, I adopt Fischer and Ravizza's (1998) guidance control as the principle underlying moral responsibility. I think that guidance control is also consistent with Johnson's third criterion for moral responsibility. Johnson simply specifies that the internal state of the agent must cause an outward event: she does not say that this cause need be a free choice between a number of alternate possible outward events and decide which to act towards.

With these weaker criteria for moral agency, some argue that animals can be moral agents. Since Darwin and (more recently) Frans de Waal (2006), animal researchers and theorists have recognized the biological foundations for morality in animals. Many times, ethologists (animal behavior scientists) describe behaviors such as altruism or cooperation among animals as *prosocial* in that they benefit not only the individual, but a larger animal collective. For example, ants collectively carrying food and wolves playing together are both prosocial.

In *Wild Justice*, Marc Bekoff and Jessica Pierce (2009) distinguish moral behaviors from "the more neutral and technical-sounding *prosocial behaviors*" (p.12). Ethologists describe behaviors such as altruism or cooperation as prosocial in order to avoid assuming that animals have morality. For example, both ants collectively carrying food and wolves playing together are prosocial. Bekoff and Pierce, however, argue that only the wolves' behaviors can be considered moral. They define morality as a "suite of other-regarding behaviors falling into the three rough clusters of cooperation, empathy, and justice" (p.138). They define requirements for a species to have morality:

1) "a level of complexity in social organization;"
2) "a certain level of neural complexity;"
3) "relatively advanced cognitive capacities;" and
4) "a high level of behavioral flexibility" (Bekoff & Pierce 2009, p.13).

On this account, certain highly social, intelligent animals act morally. Wolves meet all criteria. Moreover, when they play, they behave within strict, socially-determined rules: wolves do not bite too hard or hurt one another, lest they be chastised or bruised by the other wolves in the pack. Thus, when wolves play, Bekoff and Pierce would say this is moral behavior. Ants, on the other hand, meet criterion (1), and possibly criterion (2), depending on whether neural complexity can be considered at the level of

an entire colony; but, certainly they do not meet criteria (3) and (4); thus ants would not act morally. Bekoff and Pierce list many examples of animals acting in what they consider moral ways: elephants help injured or sick herdmates (p.84); bonobos and wolves share food (p.81, 85); rooks cooperate to gather food in controlled experiments (p.55); chimpanzees, hyenas, mice, and rats all forego food in order to save another of their species from pain (p.55-56).

The example about hyenas, rats, and macaques comes from experiments wherein two animals are put in separate cages and punished for helping one another. Here is a description of two experiments delivered on rats:

> In 1959, Russell Church demonstrated that rats would not push a lever that delivered food if doing so caused other rats to receive an electric shock. In a similar vein, in 1962, George Rice and Priscila Gainer showed that rats would help other rats in distress. In their experiment, one rat was suspended by a harness, which would cause it distress that it manifested by squeaking and wriggling. Another rat could lower the suspended rat by pressing a lever, and this is what it, in fact, did (Rowlands 2012, p.7).

Variants of these experiments were replicated among hyenas and macaques, with similar altruistic results.

These behaviors are likely motivated by empathy. In an experiment on empathy in mice, Dale Langford et al. (2006) put two mice in a cage and injected one with acetic acid to put them in extreme pain. They then observed the reactions of the other mouse. The mice observing their cage-partner writhing in pain exhibited signs of distress and pain themselves. Langford et al. concluded that mice have at least basic capacities for empathy. Bekoff and Pierce, citing the same experiments from Church (1959) and Rice and Gainer (1962), note that empathy most likely played a part in the actions of these rats to pull the lever or not. It is reasonable to think that empathy played a part in the altruistic behaviors of hyenas and macaques as well. In general, certain animals can act for the interest of others, guided by internalized moral norms and motivated by empathy. Bekoff and Pierce would say they can be moral.

Because they think animals can act morally, Bekoff and Pierce admit that animals must be able to be held morally responsible for some of their actions. However, because they argue that animals are moral agents “within the limited context of their own communities,” they argue that moral responsibility is

"species-specific and context-specific" (p.144). This is a kind of diminished moral responsibility. It precludes moral obligation in many situations for wild animals, e.g. "predatory behavior of a wolf toward an elk is amoral --it is not subject to condemnation or accolades" (p.145).

Bekoff and Pierce do not give any concrete examples of an animal acting in a way that may be blameworthy or praiseworthy. Though they concede that animals are moral agents (in a diminished sense,) it is unclear how committed they are to this idea: it seems more like an afterthought in their work, a necessary position that must be taken after arguing that animals act morally. First, they do not give any concrete examples of animals acting in a way that may be blameworthy or praiseworthy. Second, they argue against the notions of agent and patient entirely. They think these terms are "likely to promote philosophical confusion and should ultimately be avoided" (p.145).

Paul Shapiro (2006) also argues for a kind of species-relative account of moral agency in animals, wherein an animal has "responsibility only over the range of actions of which we are capable of moral understanding" (p.365). He gives examples to supplement his (and Bekoff and Pierce's) claims. If an animal decides to harm another, understanding the suffering it would cause, where the action is not needed for survival, Shapiro thinks this is blameworthy (p.365-366). For example, if a macaque in the experiments described above arbitrarily shocks another macaque (p.365). Or when a group of chimpanzees in Jane Goodall's research decided to brutally attack and kill one of their own (Shapiro 2006, p.366). Shapiro even gives examples of cross-species moral behavior that may be blameworthy. For example, Shapiro argues a dog may be blamed if she attacks her family members (p.368). Animals may also be praised. Binti Jua, an ape in the Brookfield Zoo, gently picked up and brought a 3-year-old boy to safety when he fell into the gorilla enclosure (p.367). Dolphins often hold drowning swimmers above water (p.367). Shapiro thinks these animals should be praised. Generally, if an animal can feel compassion for or understand the suffering of another species, Shapiro argues that they may be praised or blamed for their actions towards members of that species.

In *Can Animals Be Moral?*, Mark Rowlands (2012) agrees with Shapiro and Bekoff and Pierce that animals can act morally; but, Rowlands does not think that animals can be moral agents, i.e. that they

can be morally responsible for their actions. This is a striking conclusion. To my knowledge, no other thinker has argued for a distinction between the ability to act for moral reasons and the capacity to be evaluated and judged based on those actions. Thus, I dedicate the next few paragraphs to espousing his argument.

Rowlands' argues that animals can act motivated by moral emotions, i.e. emotions that have intentional content constituting moral reasons (p.35). Rowlands defines a moral emotion:

> An emotion, E, is *morally laden* if and only if (1) it is an emotion in the intentional, content-involving, sense, (2) there exists a proposition, *p*, which expresses a moral claim, and (3) if E is not misguided, then *p* is true (p.69).

To unpack this definition, in (1): what we commonly call emotions can be unintentional moods, such as sadness caused by a chemical imbalance. They can also be intentional emotions in that they are directed at something. For example, if I am angry at my brother because he hit me, my anger is directed at my brother. In (2): this anger involves a moral judgment, i.e. that my brother is wrong to hit me. In (3): Rowlands argues that if a moral judgment does not fully explain one's emotional reaction. This can be if the emotional reaction is too weak or too strong relative to its cause. For example, if my brother hit me because I hit him first, then my judgment that did something wrong is too strong; it is hypocritical. My anger is misguided. However, if my brother hits me unnecessarily, then my anger is not misguided and the moral claim that my brother is wrong to hit me is true.

Rowlands argues that we can ascribe propositional content to animal emotions (and beliefs) in human terms if there is a corresponding belief that may be expressed depending on how an animal experiences and represents the world. For example, we may say that Hugo the dog believes that “there is a squirrel in the tree,” if it is reasonable that Hugo has a corresponding belief of the squirrel and its relation to the tree in Hugo's terms, e.g. that “there is a chaseable thing up there” (p.59-61). It is reasonable to ascribe this belief to Hugo, since he can differentiate between things to chases and things to not. This avoids the problem of ascribing a belief or understanding of the human notions of ‘tree’ or ‘squirrel’ to Hugo in order to justify his belief in something that has similar content. Note that “there is a chaseable

thing up there" is only a stand-in for a possible belief of Hugo's. Since we cannot understand how an animal experiences or represents the world, we cannot express the content of an animal's beliefs. However, it is reasonable to ascribe some belief to certain animals –primarily social mammals and birds,– based on what we know of their intelligence and behaviors.

In all, Rowlands argues that an animal can have an intentional emotion that contains a moral claim which is true if the emotion is not misguided. The moral claim involved in this moral emotion can be expressed in terms of human terms if it corresponds to a belief in the animal's terms. When a moral emotion that is not misguided motivates action in an animal, the animal acts for moral reasons. All of the examples from Shapiro and Bekoff and Pierce of rats, macaques, elephants, dolphins, or any other social mammals acting out of the interest of others are motivated by moral emotions such as compassion or empathy that constitute moral reasons to act out of concern for another. Thus, Rowlands would also argue that these animals act morally. Rowlands calls these animals *moral subjects*: they can act motivated by moral reasons.

However, Rowlands does not think these animals are morally responsible for their actions; animals are not moral agents. Rowlands argues against a species- or context-specific definition of moral agency, which Shapiro and Bekoff and Pierce rely on in their claim that moral animals are moral agents. Rowlands argues that species- or context-specificity, which limits an agent's actions to the limits of their group, is "incompatible with the idea that animals are moral patients" (Rowlands 2012, p.87). Rowlands notes that if a wolf is not morally responsible for eating a deer because the deer is not in the wolf's moral context (Bekoff & Pierce 2009, p.145), then "why suppose that our behavior toward, for example, intensively raised pigs is a moral issue?" (Rowlands 2012, p.87-88). Thus, Shapiro's and Bekoff and Pierce's arguments for diminished moral agency in animals seem to be rooted in faulty notions of species- or context-specificity.

Rowlands writes that our "folk conception" of responsibility involves (1) understanding and (2) control (p.240). These correspond to what I called agent-based and mechanism-based accounts of moral responsibility, respectively. Rowlands dismisses (2), the criterion of control, as "a spurious one" (p.240).

By this, he does not mean that having control over one's actions is a spurious threshold for moral responsibility. Having control over one's actions is necessary for an action to be voluntary and blameworthy (see Aristotle on voluntary actions). I prefer Fischer and Ravizza's sense of guidance control. Rather, Rowlands means that higher-order abilities to scrutinize one's own motivations do not necessarily imply that one has regulative control over their motivations (p.170-171).

Instead of focusing on control, Rowlands defines moral agency based on criterion (1), understanding. Whereas a moral subject simply understands "*that* certain motives and actions are right" or wrong, a moral agent understands "*why* some motives and actions are right and some are wrong" (p.243). Rowlands notes that this requires some capacity to scrutinize one's own motivations and actions (p.238). Thus, I think that conceptualizing of Rowlands' criterion of understanding and Frankfurt's criterion of second-order volition (which stresses simply understanding and not the ability to deliberately choose between wills) is the best way to understand an agent-based account of moral responsibility. If a being has the capacity to reflectively endorse (to borrow Christine Korsgaard's 1996 phrase) one's own motivations, not in the sense that they can deliberately choose between wills; but, rather simply as a means of understanding why some motivations and actions are better than others, then that being is a moral agent.

In "Moral Animals and Moral Responsibility," Bert Musschenga (2015) complements Rowlands' agent-based account of moral responsibility with a mechanism-based one. Musschenga argues that animal behaviors are similar to humans' habitual actions (p.53). Musschenga applies Fischer and Ravizza's notion of guidance control to habitual actions. Musschenga argues that one can be reasons-responsive (see Section 1) via two mechanisms: "practical reason and non-deliberative habit" (p.53). Animals certainly do not have practical reason, since (as seen in the previous chapter) they cannot deliberate about their practical motivations and actions. In order to be reasons-responsive in the sense of non-deliberative habits, an actor need be able to exhibit what Musschenga calls *intervention control*, i.e. the ability to deliberately intervene in one's habitual behavior and change them based on ad hoc reasons (p.54).

While moral animals can act motivated by moral emotions (which provide moral reasons) and have some form of self-control, they do not have the kind of deliberate, reasons-based intervention control that Musschenga has in mind. Nor do they have the kind of second-order volition necessary for understanding why certain motivations and actions are right or wrong. Thus, they are not moral agents. Even moral animals are not morally responsible for their actions.

In summary, animals can be moral patients because they are sentient. This is practically ubiquitously agreed upon (see, for example, Peter Carruthers for one who disagrees). What is more controversial is whether they can act morally. Following the arguments and examples of Shapiro, Bekoff and Pierce, Rowlands, and Musschenga, I think that certain sophisticated, social animals can act morally. However, I also agree with Rowlands and Musschenga in that moral animals are not morally responsible for their actions, because they do not have second-order volition nor intervention control.

Section Three: Responsible Robots

In the section, I discussed the moral status of animals. I will now discuss the prominent claims around moral agency in machines.[14] Whereas for animals, the greatest concern is with their protection and treating them with respect, robot ethicists are primarily concerned whether robots can be blamed. I will provide an overview of the prominent arguments for and against moral agency in robots. Concluding this section, I will condense these arguments into concrete criteria for moral agency in robots. In Section 4, I will extend these criteria to animals. I will argue by reductio that the criteria for moral agency in robot ethics are too weak in that they admit absurd animals to be considered moral agents. In order to prevent such absurd conclusions, animal and robot ethicists ought to align their studies more closely around the question of moral responsibility: I will present good criteria for moral agency that is consistent across fields.

[14] Throughout this paper, I use the terms ‘machine’, ‘robot’, ‘artificial agent’ (AA), and ‘computer systems’ practically interchangeably. I prefer the term ‘machine’, since it describes a broad set of things similar to how the term ‘animals’ classifies a broad set of beings. Also, ‘robot’ has an unwanted, sci-fi connotation.

As in animal ethics, many robot ethicists argue that autonomous machines are moral subjects or moral entities but not moral agents (Johnson 2006; Sharkey 2017). However, unlike in animal ethics, this is not a universal view. There are significantly more robot ethicists who are willing to accept moral agency in machines than there are animal ethicists who accept moral agency in animals (see Floridi & Sanders 2004; Asaro 2006; Dodig-Crnkovic & Persson 2008; Hellström 2013; Floridi 2013).

Thomas Hellström (2013) introduces the notion of *autonomous power* as a criterion to assigning moral responsibility to machines. Autonomous power denotes "the amount and level of actions, interactions and decisions an agent is capable of performing on its own" (Hellström 2013, p.4) Autonomous power is in contrast to *autonomy*, which Hellström considers as environmental awareness. Hellström argues that a certain level of a combination of autonomy and autonomous power is sufficient for assigning moral responsibility to machines. Where Johnson and Sharkey argue that machines will never be moral agents because their actions cannot be separated from those of their designers, Hellström argues that advanced autonomous machines may be capable of reaching a level of autonomy and independence in their actions and how they spontaneously respond to how their actions affect their environment. He suggests that machines may be punished or praised if they are sufficiently capable of *learning* from their actions. Ultimately, he concludes that moral responsibility ought to be considered as a matter of degree, rather than a *radical dualism* (to borrow Mary Anne Warren's 1987 phrase) between humans and everything else.[15]

Hellström's point about morality as a matter of degree follows from Dodig-Crnkovic and Persson (2008) and Asaro (2006). Gordana Dodig-Crnkovic and Daniel Persson (2008) argue that moral responsibility "can best be seen as a social regulatory mechanism" which can be distributed throughout a system of interacting entities (Dodig-Crnkovic & Persson 2008, p.2). Under this view, they see "moral responsibility not as individual duty, but as *a role defined by externalist pragmatic norms of a group*" (p.2). To illustrate their definition, Dodig-Crnkovic and Persson describe maintaining a nuclear power

[15] Rowlands argues the same: that moral agents can only blamed to the extent that they understand why certain motivations and actions are right or wrong (p.240-241).

plant, where there "must be several levels of organizational and physical barriers in order to cope with different levels of severity of malfunctions" (p.3). When something malfunctions, blame is not put solely on one component of this system; it is distributed throughout multiple inter-working entities within. If intelligent systems can be included in socio-technological systems, some blame may be distributed to them. Peter Asaro (2006) argues that if a robot decides and acts based on institutional policies, then they can be praised or blamed when the institution as a whole is praised or blamed (p.14). This is how robots may be included into socio-technological systems. Both Asaro and Hellström illustrate *distributive responsibility* through militaristic and governmental organizational structures. Suppose that a soldier or bureaucrat follows orders from a superior and carries out a (morally) wrong. The blame for this wrong may be distributed between the soldier/bureaucrat and their superior. Similarly, Asaro and Hellström argue that robots ought to be considered as apart of a socio-technological system: therefore, they can be distributed some responsibility.

Since Asaro's, Hellström's, and Dodig-Crnkovic and Persson's arguments for distributive responsibility focuses entirely on how robots act within a system, they may be considered mechanism-based accounts of moral responsibility. Asaro, Hellström, and Dodig-Crnkovic and Persson rely on analogies and illustrations in their arguments; they don't give explicit criteria for which entities within a socio-technological systems would qualify for distributed moral responsibility. It seems as though they imply that any entity within a socio-technological system could be distributed moral responsibility. However, they also do not give explicit criteria as to who/what may be included in a socio-technological system. A person using a knife might be considered a socio-technological system, wherein the person (representing the 'socio-' part) uses a knife (representing the '-technological' part). Yet, if a person cuts someone with a knife, the knife is obviously not morally responsible. This is contradictory. Their notions of distributive responsibility and socio-technological systems are ambiguous. In section 4, I will devote some time to clarifying the former. For now, we can give Asaro, Hellström, and Dodig-Crnkovic and Persson the benefit of the doubt and assume that they imply that Hellström's criteria of autonomy and autonomous power are thresholds to being distributed moral responsibility.

Floridi (2013) argues that "there are clear and uncontroversial cases in which an artificial agent may qualify as a moral agent" (p.135). Similar to an account of distributive responsibility, he argues that this "does *not* relieve the creator of that agent of responsibility" (p.135). In order to be an artificial *agent* (not necessarily *morally*, but close to it,) a machine must be:

1) *interactive*, i.e. "the agent and its environment (can) act upon each other;"
2) *autonomous*, i.e. "the agent is able to change its state without direct response to interaction: it can perform internal transitions to change its state;" and
3) *adaptable*, i.e. "the agent's interactions (can) change the transition rules by which it changes state" (p.140-141).

To illustrate, Floridi gives the example of Menace, a 1960's machine that learned to play tic-tac-toe using probabilistic methods to measure the success of certain wins based on past moves and their corresponding outcomes. Viewed on the level of a sequence of games, Menace is interactive since it responds to and acts upon the state of the game. It is also autonomous, since it follows transition rules, i.e. rules that dictate how it plays the game, without being directly reprogrammed by someone else. Finally, it is adaptable, since its transition rules change based on the success of past moves. Therefore, Floridi argues Menace is an agent, because it can learn and change its rules of play between games (p.144-145). Floridi's understanding of an agent seems very close to that of a moral agent. It seems that if Menace were performing morally consequential actions, then Floridi might consider it a moral agent, at some level of abstraction.

Wendell Wallach and Colin Allen (2009) also adopt *autonomy* as a criterion for moral agency. However, they also consider *sensitivity to values* necessary (p.26). Sensitivity is a matter of degree. For example, Bruno Latour's (1999) speed bump at a crosswalk has the embodied value of slow down "so as not to endanger students:" it is sensitive to values, but to a low degree (p.186). Wallach and Allen call these sort of embodied values *operational morality*. An agent that is highly autonomous and highly sensitive to values (like humans) has moral agency. The moral capacities that autonomous machines have lies between operational morality and full moral agency, which Wallach and Allen call *functional morality* (p.26).

What constitutes being highly sensitive to values is fairly ambiguous. I think it can be interpreted in two ways (which correspond to weak and strong criteria for moral agency later). First, one can be sensitive to values without being motivated by an internal state (consisting of desires, beliefs, or other intentional states). Asaro, Dodig-Crnkovic and Persson, and Hellström would think that an artificial agent may be sensitive to values in that they are programmed to mechanistically follow good rules. In this sense, an agent need not be driven by normative obligations which can be disobeyed. An agent may blindly follow rules that a good person would consistently endorse in order to be sensitive to values in a weak sense. This is possible if a robot (with no internal state) is designed to explicitly follow good rules.

In another, stronger sense, sensitivity to values may require being motivated by an internal state. Following Johnson's criteria (1) and (3) from Section 1, this means that an agent must be motivated to follow to the best or most right action. This requires a significantly higher level of autonomy and discernment for which action is best in a certain context. This requires an intentional, internal state to motivate one's actions based on moral reasons.

Johnson argues that computer systems do not have internal states necessary for moral action. Johnson writes that

> the traditional account specifies that one of the mental states must be an intending to act. While most of the attention on this issue has focused on the requirement that the internal states be mental states, the intending to act is critically important because the intending to act arises from the agent's freedom (p.199).

Even if we could say that the complex computations that produce a machine's actions are mental states, Johnson argues that these mental states are mechanically determined: any computer that has the same programming with all the same inputs will produce the same mental state. Because their mental states are determined, machines' mental states cannot be intending to act. Thus, Johnson argues that machines cannot be moral agents. They "produce effects in the world, powerful effects on moral patients" (p.200). Hence, Johnson argues that they are *moral entities*; yet, they cannot be morally responsible. Furthermore, "those who argue for the moral agency (or potential moral agency) of computers… go wrong in viewing computer systems as independent, autonomous moral agents" (p.195).

Amanda Sharkey (2017) argues that machines are always "*human-tethered*," in that their actions cannot escape the influence of the designers and social surroundings that create them (p.212). Sharkey goes so far as to argue that we ought not to design machines that reach a threshold requirement for moral agency for fear that designers would "offload blame for mistakes or bad consequences onto robots" (p.210). Both Johnson and Sharkey argue that robots ought to be considered *moral entities* in that they make decisions that are morally consequential. Sharkey argues that robots cannot be moral agents because

1) Robots lack a biological basis for morality;
2) Robots' actions cannot be separated from the motives of their designers; and
3) Robots are not embodied: neither their body nor any others' has value to them (p.210-212).

Criterion (2) is an argument against *autonomy*. Thus, (2) is equivalent to Johnson's argument that robots lack autonomy. Criterion (3) has to do with *sociality*: robots cannot be social because they do not have a conception of any other person's body, nor the pain or suffering that they inflict on another's body. Thus, Sharkey would argue that *autonomy* and *sociality*, as well as an internalized, biological basis for morality are all necessary conditions for moral agency.

To summarize, those who argue for moral agency in current or potential machines argue that (1) *autonomy* and (2) a weak notion of *sensitivity to values* (that does not require being motivated by internal states) are necessary and sufficient conditions for moral responsibility. Call this the *weak criteria for robot moral agency*. Generally, they argue for *distributive responsibility* of machines actions in a socio-technological system. I define *autonomy* as the ability for an agent to interact with and adapt their actions in response to the environment, where their actions are not entirely determined by a designer. This definition combines Hellström's notions of *autonomy* and *autonomous power*, Floridi's notions of *autonomy*, *interactivity*, and *adaptability*, as well as Johnson and Sharkey's notions of *independence*. A weak definition of *sensitivity to values* requires that an agent explicitly follow rules to act in a way that a good person would endorse. In a socio-technological system, proponents of moral agency in robots argue that, in a socio-technological system a machine acts motivated by the moral motivations of the moral principles they are encoded with, similar to a soldier acting on commands given from a superior. In a

socio-technological system such as this, a robot may be distributed some responsibility for their actions, like a soldier being blamed for following immoral orders from a superior.

Those who argue against moral agency in machines generally argue that (1) *autonomy*, (2) a strong sense of *sensitivity to values* (which requires internal states), and (3) *sociality*. These are the *strong criteria for robot moral agency*. The definition for autonomy is the same as above. Internal states are mental states that are intending to act. Internal states are impossible if a being's mental states are determined. Thus, if a robot does not act autonomously, they cannot have internal states. A robot may be *sensitive to values* in a stronger sense if they can have intentional mental states that motivate one to act for moral reasons. For machines, *sociality* is tied to embodiment. Robots are not social because they are not embodied and do not value their own or another's body. In general, I define *sociality* as an ability to empathize with and be concerned with the well-being of other moral patients and moral agents.

Section Four: Animals and Robots

I now compare arguments between machine and animal ethics for and against moral agency. It is intuitive to think of animals and machines in the same breath. Descartes does it. They are both liminal beings, caught between inanimate objects and humans. In *Fundamental Concepts of Metaphysics*, Heidegger (1982) places animals between rocks and humans in an ontological hierarchy. Likely, he would do the same with sufficiently autonomous machines.[16] Furthermore, when considering whether or not to extend personhood, morality, or moral agency to just one of these liminal beings, we often consider the other. In "Can Intelligence Be Artificial?" Fred Dretske (1993) makes frequently appeals to animal intelligence as he argues against treating artificial intelligence as legitimate intelligence. In "When Is a Robot a Moral Agent?" John Sullins (2006) uses an extended analogy of a guide dog to consider moral responsibility in robots. Shapiro (2006) considers a gang of macaques that act empathetically in a controlled experiment; he writes that if the macaques had "been acting *robotically*… certainly we would

[16] This may be a contentious point.

have no cause to find their behavior admirable" (p.361, emphasis mine). Sharkey (2017) briefly considers the biological underpinnings of morality --drawing on Bekoff and Pierce (2009) and primatologist Frans de Waal-- in order to argue that machines do not have such social, emotional, or bodily capabilities. She makes clear that these biological foundations are not sufficient, but only necessary conditions for moral agency.

All of the above consider animals and machines only briefly. Their comments are restricted to a single paragraph or a single line. They consider these cross-comparisons only as tools for argumentation: e.g. Sharkey considers animals only to show that robots do not have the biological foundations of morality that certain animals might; Shapiro considers robot behaviors only as a thought experiment for how animals might behave mechanistically. I would like to make these connections more apparent, more central in moral considers of animals and robots. Given how intuitive these cross-comparisons are, thinkers ought to argue for criteria for moral responsibility that apply both to animals and machines. Animal ethicists ought to be reading and communicating with robot ethicists; and vice versa.

One of the only pieces of scholarship that does this is the article "Considerations about the relationship between animal and machine ethics," wherein Oliver Bendel (2016) considers moral scenarios that arise between animals, machines, and humans. Though Bendel's article sounds similar to the kind of cross-field analysis that I propose, it is not. Bendel considers "the relationship between animal and machine ethics" by way of actual interactions between machines and animals, e.g. mechanized slaughterhouses or animal-safe roadways (p.103). I propose something more like the analogies of Sullins and Sharkey, but more extensive and not metaphorical: a comparative analysis between scholarship on morality and moral agency in animals and robots. It's clear that humans are moral agents. It's clear that rocks are not. But, for liminal beings such as animals and machines, it is much more pressing that the reasons for and against attributing personhood, morality, and moral agency be consistent between these fields.

In order to maintain this consistency, I will apply the criteria for moral agency given in robot ethics to moral animals. I will argue that, under robot ethics criteria (both the criteria from proponent and

opponents to moral agency in machines,) moral animals ought to be considered moral agents. This requires translating the robot ethics criteria into definitions that can be applied to animals, then arguing that moral animals meet these criteria.

First, consider the criterion of autonomy. As stated in the previous section, many arguments for and against moral agency in robots emphasize *autonomy* as a primary threshold. I defined *autonomy* as the ability for an agent to interact with and adapt their actions in response to the environment, where their actions are not entirely determined by a designer. The last stipulation about a designer arises in response to arguments from Sharkey and Johnson that robots are not moral agents because their actions are so influenced by their design process that they cannot be considered independent entities. I would like to extend my definition of autonomy to animals. However, it is clear that this last stipulation does not apply: animals are not created by humans. So, it seems as though the criterion of autonomy (under the given definition) would too easily be achieved by animals. In order to extend this definition to animals, we must consider what undermines autonomy in their situation.

Autonomy --even in machines literature (see Hellström 2013)-- can also be defined as an understanding or independence from one's environment. Floridi (2013) notes that *autonomy* "imbues an agent with some degree of complexity and independence from its environment and from those who build the agent" (p.140). In robot ethics, it's the latter criteria that really matters: autonomy is determined by a degree of complexity and independence from those who build an artificial agent. In animal ethics, it's the former: autonomy is determined by a degree of complexity and independence from an animal's genetics and environment. This is the classical Cartesian critique of animals (and machines): that their actions are simply mechanistic responses to inputs from their environment. Scientific descriptions of animals often undermine the autonomy of their subjects by describing their behaviors as mechanistic responses that solely rely on genetically-based instinct.

This can be illustrated best by certain cases of animal constructions and environmental engineering. In *Animal Constructions and Technological Knowledge,* Ashley Shew (2017) writes that "beavers always construct their dams at the narrowest part of streams, which requires fewer materials and

less effort than building at wider stretches would" (p.92). This appears to indicate that beavers can conceive of their environment so as to be able. This seems to indicate that beavers conceive of themself and their actions independent of their environment enough to be able to to understand that a particular feature of it --namely, the width of the stream-- corresponds to the amount of work needed to build their dam. However, "observational and behavioral studies of beavers suggest that dams are built where they are because of noise volume, not because of ingenuity" (p.92). Beavers get irritated by the noise volume of the river, which is loudest at the narrowest part of the river; they build dams to cover up the sound (p.91-92). Dam-building is an involuntary response to the environment. The same is true of spiders, whose web-building behaviors are determined genetically. Spiders consistently build the same web design from when they are young to when they die: they cannot change the way they build. These animals build and adapt their environments. Yet, they do not do so autonomously. The difficulty for animal researchers to overcome when asserting that animals are agents (not necessarily *moral* agents) is to show that animals react with a level of ingenuity and not just as an involuntary environmental response. This is how animals can act autonomously.

Paul Shapiro argues that ants do not have moral agency because they act deterministically. Even though ants may behave in ways that seem outwardly moral or beneficial, their actions cannot be moral because they act deterministically. He likens ants to robots that behave based on simple algorithms informed by moral principles. In Wallach and Allen's terms, these ants have *operational morality*; yet, they do not have *functional morality* (and certainly not full moral agency.) Shapiro reasons that moral animals, unlike these ants, act in non-deterministic ways. They act autonomously.

By contrast, Shew gives a number of examples of animals that exhibit intentional technological knowledge when using tools and building constructions. Corvids (crows, jays, magpies, and ravens) can fashion twigs into complex hooks and pandanus leaves into blades for a variety of different uses (p.68-

69).[17] Dolphins place sponges on their noses and dig through the seafloor to find hidden fish (p.54). Capuchin monkeys break off flakes of stone to use as a blade (p.41). Chimpanzees fashion tools from sticks and twigs to catch ants and other food (p.39-41). Orangutans in the Milwaukee County Zoo even play with iPads! (Boyle, 3 Jan 2012). Though these are not necessarily moral actions, animal constructions and tool use show a capacity for certain animals to consider themselves and their actions distinct from their environment so much so that they can understand, manipulate, and interact with their environment in intentional, creative ways. Certainly, the actions of these animals would not be considered mechanically-determined responses to data received from their environment. These animals have the capacity to act autonomously.

Next, I consider *sensitivity to values*. In Section 3, I discussed this criterion in greater depth. There, I argued for two varietals of Wallach and Allen's definition: one with and one without internal states. Under the weaker form (without internal states,) a machine may be sensitive to values if they are programmed to explicitly follow rules which result in actions which a good person would endorse. Under the stronger form, a machine may be sensitive to values if they are internally motivated to follow moral reasons. To generalize to animals, we simply replace 'machine' with 'animals' in these definitions.

To see that moral animals fulfill both of these criteria, consider Church's (1959) and Rice and Priscilla's (1962) experiments on rats discussed in Section 2. In these experiments, rats were shown to forgo their own food in order to save another rat from being shocked or starved. They acted altruistically because they felt empathy (albeit a very simple form) for another rat when it was harmed. First, those who did not pull the lever did not simply follow their genetically-programmed desires to pull the lever and eat the food. They followed a moral emotion and chose to act for the sake of another. They acted autonomously. Second, consider that, instead of this experiment being carried out on rats, it were carried out on humans: Theo is given a lever that lets out food, but shocks Emma, if Theo pulls it. A good person

[17] Shew notes this example is especially pertinent because it shows that birds, which are usually considered lower creatures, have the capacity for intentional planning and three-dimensional spatial awareness.

would endorse Theo if he did not pull the lever. Most of the rats did not pull the lever (for similar reasons that Theo would not). Thus, a good person would endorse the rats' actions. The rats acted on a (weak) sensitivity to values.

At this point, I have shown that these rats acted autonomously and sensitive to values (in a weak sense). Thus, they fulfill the weak criteria for robot moral agency. Based on criteria from Asaro, Dodig-Crnkovic and Persson, Hellström, Floridi, and (in some sense) Wallach and Allen, these rats ought to be considered moral agents. They are morally responsible for their actions. If one of Church's rats pulls the lever and shocks another, they ought to be blamed (and possibly punished).

This is an absurd conclusion. The rats do not fully understand how good or bad their actions are, based on how much harm they inflict on the shocked rat. They have no capacity to reflectively endorse their actions or motivations. They cannot deliberately intervene to change their course of action for the better. Under the two criteria for moral agency in animals that I argued for in Section 2, namely second-order volition and intervention control, these rats come nowhere near being morally responsible for their actions. Thus, the weak criteria for robot moral agency are faulty and ought to be abandoned in favor of the criteria of second-order volition and intervention control.

An observant reader will note that the conclusion that Church's rats are moral agents is not exactly implied by the weak criteria for robot moral agency. In actuality, these thinkers argue for moral agency in a system of *distributed responsibility*. Proponents of distributed responsibility reason that the threshold for moral agency ought to be lower than traditionally construed, because they consider machines in a larger socio-technological system in which responsibility must be distributed throughout many different actors. Thus, their definition ought to be considered not necessarily grounds for full moral agency, but grounds for the distribution of responsibility within a larger social system. When an animal acts autonomously and with sufficient sensitivity to values, it can be included in a socio-animal system in which responsibility is distributed.

The problem with this view of distributed responsibility is that it is not entirely clear what kind of responsibility may be attributed to an animal acts independently of a larger socio-animal system. A rat

that pulls a lever to help a fellow rat in a controlled environment is not obviously a part of a larger system like a soldier taking orders from a superior might be. Animals' actions are often independent of a socio-animal system in which humans and animals interact and act together. However, we can simply say that an animal may act within socio-animal system that includes only themself. Thus, if an animal meets the threshold for autonomy and (weak) sensitivity to values, they may be considered moral agents in system where responsibility is distributed solely to them. The same absurd conclusions that Church's rats are moral agents follows.

Next, I consider the stronger criteria for robot moral agency. The primary differences between the weaker criteria and the stronger criteria are the additions of *internal states* to the definition of sensitivity to values and the criterion of *sociality*. Now, a moral agent must act (1) autonomously, (2) motivated by internal states that follow moral reasons, and (3) out of concern for the interests of another, especially in an embodied sense.

These three criteria are already met by Church's rats. As argued above, they acted autonomously. Their actions were motivated by empathy, which constitutes an intentional mental state motivated by moral reasons. Thus, Church's rats are sensitive to values in a strong sense. Because they are acting empathetically, they are also acting out of concern for the pain of the shocked rat. This is a concern for the body of another rat derived from a biologically-founded emotion. Thus, the actions of Church's rats indicate a level of sociality necessary to meet criterion (3). Rats may be considered moral agents under strong criteria for robot moral agency, as presented by those who argue against moral agency in robots, i.e. Johnson, Sharkey, and (marginally) Wallach and Allen.[18]

Again, this is an absurd conclusion. This is a more striking result than my argument against the weak criteria of robot moral agency, considering that these criteria are stronger and reflect the views of those who do not think that robots can be considered moral agents. These results represent a wholesale acceptance of the moral agency of animals as relatively unsophisticated as rats within the field of robot

[18] I put Wallach and Allen in both camps of robot ethicists because I included their notion of sensitivity to values as interpreted in both definitions. of moral agency.

ethics (or at least within those who have written on moral agency in machines). What I endorse as good criteria for moral agency in animals, i.e. second-order volition and intervention control, does not even consider the much more intelligent and social great apes or dolphins to be moral agents; so, it seems that the criteria for moral agency in robot ethics are too weak and permit too many animals to be full moral agents.

I think one ought to reject those criteria for robot moral agency. Of course, I leave myself open to the possibility that I have not read every argument for moral agency in robot ethics (though I have attempted to make a valiant effort to demonstrate the key voices here). If there are thinkers that do not argue in the way that the robot ethicists I have characterized here do, then my argument may simply not extend to them.

Regardless, I think that the criteria of second-order volition, in the sense that it allows an agent to understand why their motivations and actions are good or bad, and intervention control, which allows an agent to deliberately intervene and change their habitual actions for the better, are well argued-for and are based in strong support from well-developed moral theories around moral responsibility. Even if I cannot show that a thinker's work falls into my characterization of the current state of the field of robot ethics and leads to an absurd permittance of rats as moral agents, I still think there are good reasons to prefer second-order volition and intervention control as thresholds for moral agency.

Furthermore, I think these criteria for moral agency ought to be consistent across the fields of animal ethics and robot ethics. Animal ethicists mention machines in their arguments and robot ethicists mention animals in their arguments all the time. (See the beginning of Section 4 for a list of examples.) When they do so, they implicitly reason that animals and robots are similar beings for which it is fruitful to compare to one another. We can understand more about animals by thinking about robots. And vice versa. Ultimately, we can better understand humans and our moral nature by comparison with these liminal beings.

Rather than maintaining distinct fields of animal and robot ethics, which ask similar questions of their subjects and consider other subjects only as props for the sake of argument, we ought to explicitly

bring out the similarities and dissimilarities between animals and robots. There are ways in which robots are not like animals. For example, robots are built and animals (save pets and domesticated animals, in some sense) are not. This has some consequence for the moral status of robots; and for the moral status of animals. The only to understand what this consequence is is by directly comparing the two (as John Sullins does).

However, there is no reason that animals and robots are so dissimilar that they deserve entirely distinct definitions of moral agency. Animals do not have the capacity second-order volition nor intervention control. Nor do robots. Thus, I do not think that either are moral agents. Most adult humans have both capacities; thus, they are moral agents. If future robots meet these thresholds, they might be considered moral agents. One definition applies to all three subjects --humans, animals, and robots. Animal and robot ethics ought to hold consistent definitions of moral agency. Or, at the very least, robot ethics ought not to have definitions of moral agency which are so faulty that they admit rats to be morally responsible for their actions.

To illustrate my point further, it would only take a reading of Mark Rowlands' definition of a *minimal* moral subject to recognize that Wallach and Allen's notion of sensitivity to values, even motivated by an internal state, is too weak of a notion for moral responsibility:

> X is a *moral subject* if X possesses (1) a sensitivity to the good- or bad-making features of situations, where (2) this sensitivity can be normatively assessed, and (3) is grounded in the operations of a reliable mechanism (Rowlands 2012, p.230).

Here, criteria (1) and (2) in Rowlands' definition of a moral subject look remarkably similar to Wallach and Allen's notion of sensitivity to values in their definition of moral agency. Obviously, Rowlands' definition for a moral subject is much too weak to be considered as a good definition of moral agency. Wallach and Allen's definition ought to be revised.[19]

[19] I recognize that Wallach and Allen could not have foreseen this, since *Moral Machines* was published in 2009, while Rowlands published *Can Animals Be Moral?* in 2012. However, the notion of sensitivity to values in robot ethics does not seem to have been revised yet.

I anticipate an objection here. I am arguing that, in terms of moral agency, there ought to be a consistent definition that applies to both animals and robots alike. Yet, as moral patients, animals and robots are very dissimilar. Animals are sentient and they may suffer. This is why they may be moral patients. Robots may never be truly sentient and may never be designed to suffer (or even have emotions). However, at some point, we might imagine a highly intelligent, relatively social robot that it might feel intuitively wrong to destroy or harm.[20] In this case, the definition for a moral patient in animals (or any living thing) seem as though they need to be different from the definition of a moral patient in machines.

Even though animals and robots are dissimilar, and may always be dissimilar, in the case of sentience and whether they may be considered moral patients, I think that we can offer an expanded definition of a moral patient that encompasses both robots and animals. If it is true that we will feel an intuitive sense of wrong being done to a highly intelligent robot that lacks sentience or suffering, then this reveals that sentience and suffering alone are not adequate for understanding who or what may be a moral patient. To understand these intuitions of wrongdoing, our definition of a moral patient would need to be expanded to consider a high level of intelligence as grounds for a being to be on the receiving end of moral actions. This would apply not only to highly intelligent robots, but to animals as well. Imagine an animal that has no emotions and simply follows determined rules, but does so in a way that indicates an high level of intelligence (one might imagine a collectively-intelligent ant colony or even a network of trees and fungi). If a high level of intelligence were grounds for considering a robot to be a moral patient, then there is no reason why a similarly intelligent animal would not be a moral patient as well. In sum, definitions of moral status --be it that of a moral patient, subject, or agent-- hold across kinds of beings. If traits in one beings indicate a need to change our definition, then this ought to apply to all other beings as well.

[20] I do not believe this. It is simply a hypothetical.

Conclusion

The aim of this chapter was to present arguments for and against moral responsibility in animals and robots and condense them into a well-formed definition of moral agency. I began by touching on some of the most influential considerations of personhood and moral responsibility from Frankfurt and Fischer and Ravizza, as well as a concise synthesis of classical arguments by Deborah Johnson. I then argued, following Regan, that animals are moral patients because they are sentient. I outlined arguments for and against moral action and moral responsibility in animals. I prefer Mark Rowlands' understanding of the moral lives of animals. Certain animals --largely social mammals and birds-- may be considered moral subjects in that they can act morally via moral emotions. However, they are still not moral agents. To be a moral agent, one needs the capacity to reflectively endorse or understand their own motivations and actions as good or bad and to deliberately intervene to alter one's habitual actions. Because animals --even moral animals,-- do not have second-order volition nor intervention control, they are not morally responsible for their actions.

Following a discussion of animals, I presented a reading of the question of moral agency within robot ethics. I found the field largely separated between those who argue for and those who argue against moral agency in robots. I synthesized many of these thinkers' arguments into the following criteria for robot moral agency: (1) autonomy, (2) sensitivity to values, and (3) sociality. (1) and (2) make up what I call the weak criteria for robot moral agency. Adding (3) makes up the strong criteria. I applied these definitions to moral animals, arguing for an absurd conclusion in both the weak and strong criteria for robot moral agency. I argued for a rejection of these criteria in favor of the definition of moral agency, which relies on reflective endorsement and intervention control. Because robots --even autonomous robots,-- have neither of these two capacities, they are not (currently) morally responsible for their actions.[21]

[21] These characterizations of autonomous robots and moral animals as amoral agent may change in the future, if significant research into either area reveals a capacity for reflection or guidance control.

Finally, I argued that this kind of argumentation via explicit comparison between animal and robot ethics ought to be considered. These two fields ask similar questions about similar liminal beings. And even where animals are not similar, direct comparison reveals interesting dissimilarities. These kinds of cross-field comparisons will, in turn, reveal more about humans and our moral nature.

Chapter 3: Building and Moral Responsibility

The discussion of moral responsibility in animals stemmed from a consideration of animality in Steven Vogel's work, in Chapter One. Vogel's central thesis of postnaturalist environmentalism is that building implies responsibility. I presented principles, by way of Heidegger and Marx, for demarcating between humans and animals. Drawing on Frankfurt, who argues that personhood and moral responsibility are tied at the hip, I used Heidegger's and Marx's meditations on animality to provide definitions. Based on these definitions, I argued that humans are solely morally responsible for what they build. In Chapter Two, I presented how these ideas fit into a larger study of the moral status of animals. I gave arguments as to why animals are not morally responsible for their actions and argued for consistency in the principles underlying moral agency in humans versus liminal beings, i.e. animals, robots (we might include children here as well).

One will notice that there is a slight disconnect between these two chapters. The first chapter is about moral responsibility for building; the second is about moral responsibility for action. Building is a kind of action. Thus, one might be inclined to extend the results from Chapter Two on moral agency to building: animals are not morally responsible for their actions, thus they must not be morally responsible for what they build. This is true. However, it does not mean that the capacity to be morally responsible for one's actions is equivalent to the capacity to be morally responsible for what one builds.

A parent, in some sense, builds their child; however, they are not morally responsible for their child's action once she is a rational adult. I do not think we would say that the actual builders, the construction workers, that built the City Center Mall in Columbus, OH are responsible for the failure of that mall. By contrast, parents are responsible for the actions of their young children. Engineers are responsible for the things they design. Oppenheimer --who led the team that built the atomic bomb-- is responsible (in part) the death and destruction in Hiroshima and Nagasaki. Dog owners are responsible if their dog bites.

These examples of building are vast and seemingly dissimilar. Vogel's understanding of building is no less nebulus: he includes dams, machines, toasters, skyscrapers, a hut in the Black Forest, the Dow

Jones Industrial Average, shopping malls, and the environment. In this chapter, I will clarify what building is in a way that encompasses all of these examples. I will also discuss how we are (and are not) morally responsible for what we build.

I will begin by discussing how moral responsibility arises in engineering ethics and philosophy of technology. I will then give a definition of building informed by Vogel, Marx, and Heidegger. I will then discuss how humans can be responsible for autonomous beings they build, drawing on literature about pets and children. Finally, I will discuss problems of present and future autonomous machines, drawing on the field of ethics of algorithms.

Section One: Vogel on building

Steven Vogel (2015) considers building a social, material practice. The paradigm case of building for Vogel is constructing a mall. This is social in that an entire construction crew needs to work together in order to lay down a foundation, put up a frame, the walls, the floors. the ceilings. As well, all the materials --the concrete, the steel, the glass,-- are all produced and shipped by others. Especially in our increasingly global economy, building involves a massively social network of workers, engineers, designers, planners, etc. I agree with this. In general, I quite favor Vogel's definition.

However, I think Vogel's definition of building lacks in two ways: (1) it does not account for how one may be morally responsible for what they build; and (2), Vogel's definition simply does not encompass all the things that Vogel himself uses as examples of building. There are things which Vogel says are forms of building that do not fit under his definition of building as material construction.

First, Vogel's central thesis is that building implies responsibility. This is an intuitive notion; not one to argue against. However, Vogel seems to make exceptions to this rule without adequately explaining why they are exceptions. For example, he argues that we are not directly morally responsible for the environmentally harmful actions we need to do on a daily basis, e.g. when Vogel drives to work in the morning because he cannot walk, bike, or take the bus. Vogel argues that, since we have no control over how our roads and cities are structured and we do not understand their structure in terms of

environmental actions, we cannot be morally responsible for, say, driving to work in the morning. However, we are responsible in that we have the capacity to collectively organize and democratically restructure our society (here, he means our society's material structure) in order to encourage more environmentally-friendly individual actions. In Chapter One, I argued the same point from the standpoint of Marx's understanding of animality and alienation. I extended Vogel's argument to imply that those who are closer to positions of power in determining the structure of a society or economy are most responsible for how we build the environment. This is a very different, much more complicated thesis than simply that 'building implies moral responsibility'. A consequence of this, which I argue that Vogel would condone, is that there are certain persons (e.g. marginalized or disenfranchised persons) would not be morally responsible for their individual environmental actions if they do not have any say in how their society is organized. This is a case where someone builds the environment through their individual actions, yet they are not morally responsible. This is an exception to the rule that building implies moral responsibility. Vogel's definition of building does not make immediately evident why this is an exception.

Consider another example of an exception to the thesis that building implies moral responsibility that cannot be fully explained by Vogel's definition of building. Workers built the highway system. They built the streets that the buses run down. They even built the buses. Vogel drives to work in the morning because the neighborhood he lives in is too far from his place of work to walk, bike, or take the bus. His car emits greenhouse gases. He would emit far less if he walked, biked, or took public transit. In this case, Vogel is not to blame for the architecture of his city and its lack of sustainable means of transportation. Nor are the workers who laid the concrete that made the highway or forged the parts that made the bus. They did so as a social practice. They created the material conditions of the city, namely the highways and the streets and the bus. The workers built the city (following Vogel's definition of building to a T). However, they should not be held morally responsible for the environmentally-poor design that forces

Vogel (and everyone else) to emit more greenhouse gases on his drive to work.[22] It is not those who engage in the material practice of building that are responsible for what they build.

Those who planned the city, those who designed its overall structure, should be morally responsible. The city planners and politicians and car company executives and lobbyists are really the ones who built the city. They did not do so materially: they were not driving the bulldozers and backhoes, laying the pavement. They did so immaterially. They laid the ideal foundations for those workers (mentioned above) to follow their plans and build the material structure of the city.

The distinction between material and immaterial forms of building arises out of Marx. In the *Grundrisse*, Marx [1858] adopts the classical political economists' (Adam Smith and the like) categories of productive and unproductive labor. Productive labor "is only that which produces *capital*," i.e. a material thing which has value ([1858], p.302).[23] In *Provincializing Europe*, Dipesh Chakrabarty (2000) critiques Marx's notions of productive and unproductive labor via Marx's example of the piano maker and the piano player. Marx rhetorically questions whether both the piano maker and the piano player are productive workers, since "the piano would be absurd without the piano player" ([1858], p.302). Marx answers himself, 'no': the pianist may "produce music and satisfy our musical ear," and even "to a certain extent produce the latter [our musical ear];" yet, making music is not "*productive labour* in the *economic sense*" (p.302).[24] When the pianist makes music, Marx writes, it is no more productive than "the labour of the madman who produces delusions" (p.302). Marx acknowledges a de facto distinction between the

---

[22] I may accept that they can be responsible in an implicit sense. In an article published in *Science & Engineering Ethics*, Eric Katz (2010) describes how designers of crematoria ovens --though they were not told of their uses-- engineered the ovens with questionable characteristics, e.g. to permit "high capacity and more efficient burns in the ovens" (p.575). Though they acted under the orders of higher-up SS officials, these engineers may be complicit in the death of millions in concentration camps, because they should have had a clue into what they were designing. Similarly, if a bridge construction crew should have reason to be suspicious of the bridge they build, they may be complicit and somewhat morally responsible for it.

[23] This distinction between activities that produce capital and those that do not reminds me of Aristotle's notion of *poiesis*, which I will discuss shortly.

[24] In a very interesting analysis that is, sadly, totally unrelated to my current argument, Chakrabarty notes that this is "the closest that Marx would ever come to showing a Heideggerian intuition about human beings and their relation to tools. He acknowledges that our musical ear is satisfied by the music that the pianist produces. He even goes a step further in saying that the pianist's music actually --and 'to a certain extent'-- 'produces' that ear as well" (p.68).

labor of the piano maker who literally constructs the material piano versus that of the pianist and the madman who ideates. Then, “Marx both acknowledges and in the same breath casts aside as irrelevant the activity that produces music” (Chakrabarty 2000, p.69).

Chakrabarty critiques Marx’s dismissal of the pianist’s labor, since it constitutes a dismissal of an important history of labor (p.69). In “Immaterial Labor” Maurizio Lazzarato (1996) introduced the term *immaterial labor* in part to recognize activities of the artist or the influencer of public opinion, which have been dismissed as “work” (p.132).

The primary reason that Lazzarato introduces *immaterial labor*, however, is to define the “informational content of the commodity,” i.e. changes (circa 1996) in the workplace wherein “skills involved in direct labor [became] increasingly skills involving cybernetics and computer control” (p.132). To give a formal definition, immaterial labor is “labor that produces an immaterial product, such as ideas, images, forms of communication, affects, or social relationships” (Hardt 2005, p.176). “It is not primarily about making a material object, like the work that makes a car roll off an assembly line or extracts coal from a mine;” to this, Vogel would add work that builds a mall (Dyer-Witheford & de Peuter 2009, p.4).

In *Empire*, Michael Hardt and Antonio Negri (2000) describe immaterial labor as “the hegemonic form of labor” in today’s day and age. They do not mean that immaterial labor is dominant in that the most number of workers produce immaterially; but, rather that capital depends on computers or creative design to lead and plan out most of the work that is being done today. Nick Dyer-Witheford and Greig de Peuter (2009) identify the videogames industry as a site of an enormous wealth of immaterial labor. For Dyer-Witheford (2015), recognizing immaterial labor as a form of labor is a way of drawing connections between the newly-emerging digital managerial class, e.g. IT workers, engineers, architects, and the material labor of the classical proletariat class, e.g. those in the Global South who mine for cobalt or work in a factory assembling electronics. Immaterial labor is a theoretically and practically important and well-established notion that is distinct from the notion of material labor. Both forms of work rely on one another. But, for the sake of this chapter, I draw on this distinction between immaterial and material labor for the same reasons that Dyer-Witheford (2015) does: to recognize the work of engineers, designers, and

architects in the unique ways that they build. The immaterial work of planning and design is emerging as the most morally consequential form of building today.

Vogel's definition of building is as a material act. He does not recognize immaterial labor as a form of building. He claims that building implies responsibility; yet, those who build materially are rarely responsible for what they build. More often, it is the planners, the architects, the designers, the engineers that are responsible for how a thing is built. Because Vogel considers building to be strictly a material practice, he precludes considering how one may be morally responsible for their immaterial labor in the building process.

Second, Vogel's philosophy of building suffers from an internal inconsistency in that his definition of building does not encompass all of the things that he lists as examples of building. For examples of things that we build, Vogel lists dams (p.23), tools (p.84), machines (p.151), toasters, skyscrapers, a hut in the Black Forest (p.169), the Dow Jones Industrial Average (p.78), and, most importantly, shopping malls (p.131) and the environment (p.30). These are all (save the possibility of the environment, which I will explain shortly) constructed.

Usually we think of building as construction. We build a house or a bridge or a mall. To build colloquially means to construct an edifice. Marx recognizes production as the fashioning of natural materials into human means --construction (see the *1844 Manuscripts* (1975) and *Capital* [1867], Volume One, Part 3, Chapter VII, Sections 1 and 2). Vogel also considers building as construction: as social construction (Vogel 2015). Vogel takes the notion of social construction literally. Usually when we invoke social construction, it is to reveal something as contingent upon our history and social organization, and therefore mutable (p.36). For example, if we argue that the notion of race is socially constructed, it is to say that the idea of race was created in a genealogy of social contexts. If the categories of race lead to racial discrimination (as in the United States,) then this idea ought to be changed. Revealing race as a social construction is recognition of the idea of race as contingent upon our societal and cultural structure, which can be changed intentionally by its members. Vogel does not mean social construction in this sense. Rather, Vogel argues that, as a society, we literally construct our environment.

Vogel understands building as social, material construction.

In addition to the mall, toasters, and the environment, Vogel also lists ecological restorations (p.111), babies, and purebred dogs (p.151) as built things. These are not constructed. To show why, I will focus on the example of ecological restoration.[25] Restoring something is not the same as physically constructing it. The builder of a forest does not place trees into the ground and nail leaves onto the branches. The process of building a forest relies on caring for and implicitly shaping the trees and plants as they grow, like shaping a bonsai tree.

In his defense, Vogel's definition of building as material construction grounds some very important conclusions in his philosophy (which I agree with).[26] For one, Vogel argues that every single action that we do, in some way, transforms and constructs the larger material structure of our environment (p.56). This allows us to understand our individual actions in how they contribute to the larger collective formation of the environment. This aligns with the environmental virtue of "self-knowledge" (p.117). For another, material things are hard to change. We often construct things poorly (p.119) and the things that we build escape our intentions once we bring them into the material world. Thus, our practices ought to be "marked by a spirit of tentativeness and fallibilism, not by hubris" (p.166). We ought not think of ourselves as masters of the world who can construct the environment exactly as we want to (p.167). This aligns with the environmental virtue of "humility" (p.119). In combination, Vogel writes that

> self-knowledge and humility might be the key environmental virtues, teaching us of our responsibility for the world we inhabit on the one hand but also reminding us not to overestimate our ability to remake it in any way we want on the other (p.121).

These environmental virtues are a direct consequence of Vogel's understanding of building as material construction. Yet, I do not think this restricted understanding of building is the only way to argue for the importance of these virtues. This is the aim of this chapter: to show how and why we are responsible for what we build. This is a form of self-knowledge. It also allows us to understand how we

[25] Though I agree with Vogel in that babies may be built, in a sense that I will discuss later, I recognize that this may strike many as odd examples. Eric Katz (1992) also discusses bred dogs and ecological restorations as *living artifacts*.

[26] I think his definition of building as social is fine. I see problems with building as material construction.

build poorly --and thus how we ought to be humble-- when we build in immaterial ways or ways that are not encapsulated under the name ‘construction’. I think that we can give a definition of building that consistently describes the numerous examples of building that Vogel describes and demonstrates how we may be responsible for what we build.

Section Two: Building as Cultivation

In “Building Dwelling Thinking,” Heidegger (1971) makes a distinction between construction and cultivation. He writes: “Building as dwelling unfolds into the building that cultivates growing things and the building that erects buildings” (p.146). The primary case of construction is raising an edifice (p.145). The building that cultivates growing things is what restoring a forest, raising a domesticated animal, or caring for a baby is like.

Even in Heidegger’s work, the paradigm case of building is construction: building a bridge (p.145). He writes that cultivation, i.e. “[b]uilding in the sense of preserving and nurturing is not making anything” (p.145). I read this as saying that cultivation allows something to grow and make itself; one does not directly construct it. Heidegger uses the example of when a farmer tills her soil or cultivates her grape vines (p.145).

Cultivation is extremely important for Heidegger. As discussed in my Chapter One, Vogel and Heidegger disagree about the materiality of building. Heidegger considers the hand (which builds) primarily as a hand that writes. For Heidegger, building is rooted in thinking. For Vogel, thinking is rooted in building. Heidegger considers writing a form of building. Particularly, he considers writing cultivation. Byung-Chul Han (2017) notes that, for Heidegger, writing is “a matter of cultivating, plowing, and tilling language like the soil” (p.38). This is an interesting, important point in Heidegger that I must but briefly gloss over for the sake of my argument.

What is important for my argument is that Heidegger considers writing, an immaterial act, a form of building. In particular, it is a form of cultivation. Thus, cultivation can be material or immaterial. Construction is strictly material. Writing is closer to the work of the engineer or the designer than it is to

the work of the construction worker. The civil engineer who designs a bridge does not physically build the bridge. She relies on the construction workers to do so for her as she shapes the project. This is remarkably similar in form to the shaping of a bonsai tree or an ecologically-restored forest. The civil engineer allows the bridge to grow on the backs of the construction workers. Thus, I argue that the planning that engineers, architects, or designers do is a form of cultivation.

Considering cultivation as a form of building allows us to consider ecological restorations, dogs, and babies as built things. It also allows us to consider how engineers, designers, and other planners build in immaterial ways and how they may be responsible for what they build.

It also allows us to understand how we build the environment. In many ways, we do not directly construct the environment. We let off pollutants like DDT or greenhouse gases that stifle the growth of many of the lifeforms that make up our environment. Our responsibility, then, is to care for the environment by cultivating an environment which can grow into a beautiful, amiable, good form. Cultivation is associated with the environmental virtues of care and humility (which Vogel would approve of). Considering the environment as built through cultivation allows us to understand the ways in which we currently do not care, and the ways in which we ought to care, about the environment.

Next, I will introduce a definition of building that encompasses both construction and cultivation. I will then consider why and when we are morally responsible for what we build.

Section Three: What is building?

Building brings something into being. For example, when a carpenter builds a house, they bring the house into being. This definition is simple. This makes it wider than the common understanding of building, which largely aligns with Vogel's definition of building as construction. Colloquially, building has the connotation of constructing a non-living thing out of non-living materials. For example, we would say that a carpenter builds a house, but a gardener grows a garden. If the gardener plants and cares for the plants, she is bringing it into being as much as the carpenter does a house. Thus, under my definition (which includes cultivation,) a gardener builds a garden.

This definition looks strikingly similar to Aristotle's notion of *poiesis*: to bring something into being that did not exist before. Cultivation brings something into being; though, it is not prima facie clear that it brings something into being which did not exist before. In fact, Heidegger does not think it does: he writes that cultivation i.e. "[b]uilding in the sense of preserving and nurturing is not making anything" (1971, p.145). When a gardener cultivates a garden, they are not bringing the plants themselves into being. The gardener is not making something, like the carpenter who constructs a house.

However, I think that the gardener still brings a garden into being in the sense that she cultivates the plants and gives them a form. Suppose that a patch of plants grows naturally, growing into a ragged, scattered patch of plants. The gardener weeds and trims in order to bring out the form that of a well-kempt garden out of this patch of plants. Through cultivation, the gardener forms how the plants grow and how they are formed. In this sense, I think that the gardener brings a garden into being. Thus, cultivation brings something into being in that it presents a growing thing in a certain form. In this sense, I think cultivation is like *poiesis*.

However, *poiesis* connotes an abstract creative process, separate from practice. I do not wish to make this separation for building. Construction brings something into being as a material practice. Cultivation may also bring a material thing into being. However, cultivation can also be an abstract creative process, like engineering, design, or architecture. These activities are closely related to *techne* as well, since building is a technological activity. In all, I do not mean to limit building to a specific kind of bringing something into being, e.g. strictly material or immaterial, or strictly *poiesis*. My definition should open to a wide variety of forms of building, which encompasses many of the examples Vogel lists. However, my definition should be understood semi-informally: building ought to be understood in light of its common sense, technological connotations.

The aim of this chapter is to understand how and why, fundamentally, we are morally responsible for the things we build. With a definition of building established, the next section is dedicated to understanding building as an action that we may be morally responsible for.

Section Four: Building and moral responsibility

Building is an action. Like any action, building has a direct effect. A carpenter builds a house (the action) and a house is built (the effect). Yet, building is different from regular actions in that a consequence of the act of building may be an action later enacted through or by a built thing. For example, when a roofer hammers a shingle into a roof, there is the direct action that drives a nail into the roof through the shingle. However, there is also a later effect wherein water flows over the top of shingle. Building not only causes the direct effect of a thing being built, but it also causes future (indirect) effects enacted *through* or *by* the built thing. The rainwater runs off the roof; this is an effect enacted *through* the characteristics of the shingles (the built thing). A house creaks; this is an effect enacted *by* the house (the built thing). This latter form of indirect causation is characteristic of building. Building causes deferred effects, which depend on the built thing.

Deferred effects that building causes are interesting when considering moral responsibility. How one can be morally responsible for the direct effects of an action is an old and saturated problem stretching back at least to Aristotle. Very few (if any) have discussed why we are responsible for what we build. For example, suppose that Gerry intentionally pours a bucket of water into Mary's attic. This is a direct action that causes Mary's attic to flood. Gerry should be blamed for any grief it causes her. On the other hand, suppose that Gerry intentionally builds a faulty roof that lets rainwater into Mary's attic. This is a deferred effect. Gerry should also be blamed for her grief. In the former situation, Gerry let water into the attic. In the latter situation, Gerry built the roof. The roof let water into the attic. The latter action is not direct. If responsibility were assigned like any other direct action, the roof would be responsible for letting water into Mary's attic. Intuitively, however, we know that Gerry is responsible for the leakage. It is only through the faulty roof that this effect is enacted. The study of moral responsibility involving normal, direct actions does not immediately apply when considering deferred effects from building. This is what I focus on in this chapter: why is a builder morally responsible for the deferred effects of the thing they build? In other words, why are we morally responsible for what we build? For example, why is Gerry morally responsible for letting water into Mary's attic?

Furthermore, building may not be viewed as a string of direct actions and effects that cause each other. If Joe knocks down a domino that hits another domino and another and so on, Joe is responsible for knocking down not only the first domino but every domino after it. By contrast, when Gerry builds a roof, it is not a direct consequence that water will leak. It would not leak if it simply did not rain. The rain, in some sense, causes the leakage via the faulty roof. This effect (the leakage) is indirectly caused by the faulty roof, and thus by Gerry. When Gerry builds a faulty roof, there is a stored potential that the roof will let water into the attic. Because of this, one cannot simply say that Gerry is responsible for letting water into Mary's attic like Joe is responsible for knocking down the last domino in a line.[27]

In fact, I do not think that the kind of responsibility that I am proposing --responsibility for the deferred effects of a built thing-- can be described by any other major framework of moral responsibility. While they are very similar, theories of individual (mentioned at length in Chapter 2), collective (see Feinberg and Arendt), and distributed responsibility (see Asaro (2006), Dodig-Crnkovic & Persson (2008), Hellström (2013), and Floridi (2013)) cannot entirely encompass why and how we are morally responsible for what we build. Individual moral responsibility looks like

Agent → Patient,

collective responsibility looks like

Collective → Patient,

distributed responsibility looks like

Collective → Individual in collective system → Patient,

and responsibility through building looks like

Agent (Builder) → Built Thing → Patient.

All of these four schematics look different and are unique in some way. They have similarities, but cannot be encompassed entirely by one another. Thus, building implies a unique form of moral responsibility that needs its own theory to be properly understood. I rely on similarities between these other forms of moral

---

[27] Here I use 'responsible' rather loosely. The same point about Joe's dominoes and causal responsibility may easily be made about moral responsibility.

responsibility and draw insights from them to illustrate how and why we are morally responsible for what we build.

Though not directly addressed in a unified theory, much of philosophy of technology is dedicated to understanding how engineers and designers are responsible for what they build. From Don Ihde's notions of embodiment relations and multistability to Bruno Latour's notion of script, philosophy of technology illuminates how a technology communicates these embedded values from the point of view of the user of these technologies. Postphenomenology understands how technologies determine how we experience (with) the world. As a value-neutral example, looking through a pair of glasses, the wearer "experiences a transformed world" (Rosenberger 2017, p.70). There are less neutral examples. Peter-Paul Verbeek illustrates how a sonogram transforms one's world in an "ethically non-innocent manner" (Rosenberger 2017, p.71). The fetus, in actuality, is small and engrossed by fluid and flesh with the womb. Yet, the sonogram zooms in on the image of the baby, making it seem larger and more separated from its mother's womb –which now presents itself only as an environment to house the fetus. The sonogram shapes how one sees a baby in a womb; this influences mothers to treat their fetus as an independent child when considering medical intervention, e.g. having an abortion or a C-section. In this sense, the designers of the sonogram embed certain values into the technology. Their design choices influence how a mother makes moral decisions based on how they see their baby through the sonogram. In general, designers embed values into the things they build by shaping the way that the technology shapes our experience of the world.

Wallach and Allen (2009) write that "[o]ne of the major accomplishments in the field of 'engineering ethics' over the past twenty-five years has been the raising of engineers' awareness of the way their own values influence the design process and their sensitivity to the values of others during it" (p.25). A famous example of gender bias in engineering is that seatbelts are far more dangerous to women because cars are often designed with only a crash test dummy the size and build of an average male (Shaver 25 March 2012). Because they created these faulty seatbelts, these engineers might be considered responsible for the harm that was caused by it (I will consider when we may and may not ascribe blame to

these engineers later).

This is also true of engineering in an algorithmic age. Algorithmic technologies automate human actions. This makes them unique in that many algorithmic technologies do not need a human actor to make decisions or act. In this sense, algorithmic technologies may be autonomous. Yet, we still consider software engineers responsible for their algorithm's decisions. Often in computer science and engineering, algorithms are presented as mathematical constructs (see Hill 2016), which are "apolitical, technocratic, and value-free" (Danaher 2015, p.5). The field of ethics of algorithms approaches algorithms from the perspective of how they are understood by the general public (Mittelstadt et al. 2016, p.2). The public understands algorithms as embedded in a specific technology, often one which makes socially-consequential decisions, e.g. search for information or assign a credit score (see Kraemer et al. 2010; Danaher 2015; Mittelstadt et al. 2016). Designers create these algorithms with embedded values so that, when they make socially-consequential decisions, they acquire an ethical dimension (Kraemer et al. 2010; Danaher 2015; Mittelstadt 2016). Often, this leads to algorithms that perpetuate discriminatory decisions based on gender, race, class, etc. (for popular discussions of algorithmic bias, see O'Neill 2016; Noble 2018; or the Algorithmic Justice League). For example, early optical character recognition (OCR) algorithms could not recognize black faces or search algorithms associate the word 'man' with 'doctor' and 'woman' with 'nurse'. Software engineers embed values into the algorithmic technologies they make in that these technologies make socially-consequential decisions with an ethical dimension. Often, this plays out in the form of perpetuating sociodemographic biases.[28]

In an article on engineering and responsibility written for *The Ethics of Technology*, Jessica Nihlén Fahlquist (2017) makes a couple distinctions important for understanding how engineers are responsible for what they build. Nihlén Fahlquist distinguishes between backward-looking and forward-

[28] There are other values embedded in algorithms. Many algorithms are embedded with values to disregard a reasonable expectation of user privacy (Nissenbaum 2010). As well, social media algorithms prefer content that is more sympathetic to a user's political leanings and rarely exposes users to content that will challenge their worldview, leading to undemocratic degeneration of the ability to dialogue across difference (Sunstein, 2006; Pariser, 2011; Bozdag, 2013).

looking responsibility (p.133). Backward-looking responsibility is what I have been referring to (and will refer to) generally as moral responsibility or blameworthiness. It is the capacity to look back on an action or event and blame someone for it in a way that is fair. Forward-looking responsibility, on the other hand, has an "efficacy aim" to prevent an action or event in the future (p.133). Ascribing forward-looking responsibility to someone is to hold them responsible, not necessarily to uphold standards of fairness and morality, but simply for preventing an action or event in the future. We often ascribe backward-looking and forward-looking responsibility to the same person, they are not the same. However, there are instances where we would say one is responsible in a forward-looking sense, but not in a backward-looking sense.

This is often true for engineers and designers. Forward-looking responsibility is particularly important for engineers who deal with "technological risks" (p.133). For example, a mother was hit and killed crossing the street on Maryland Avenue on the East Side of St. Paul, MN in 2016 (Gottfried May 24 2016). It was a grave accident; if anybody, the driver is to blame. Yet, the task of preventing future pedestrian deaths on the street was charged to the civil engineers at the city. They restricted the number of lanes from two to one at crosswalks and put in large cement medians intermittently in the middle of the street to slow drivers down. The driver who hit the mother might be ascribed backward-looking blame. The civil engineers were not: they were ascribed forward-looking responsibility because they have the power and means to change the road in order to prevent future deaths. Engineers and designers, in general, have the capacity to change our material world to promote or prevent future outcomes. They are often ascribed forward-looking responsibility only.

However, in many of the other examples listed above, e.g. Gerry's roof that leaks rainwater into Mary's attic, seatbelts that disproportionately harm pregnant woman, or OCR algorithms that do not recognize black faces, it seems reasonable to blame the designers or builders. The project of assigning moral responsibility to a builder for the thing they build must be wary of this double meaning of responsibility --as forward- or backward-looking-- and be able to decipher when a builder is blameworthy and when they are only responsible in a forward-looking sense.

Section Five: When are we not responsible for what we build?

As discussed earlier, civil engineering and architecture may be understood as forms of cultivation, since the designer writes and plans the form of a built thing and the construction crew constructs the built thing into this form --similar to a gardener planning the form of the garden and relying on the plants to grow into this form. If all the plants in the garden grow, the flowers bloom, and the garden still looks ugly, then this is the gardener's fault. The plants did as they were supposed to, growing into the form that the gardener supplied. The gardener planned an ugly garden. Similarly, we would not blame the construction crew for the collapse of the Florida International University (FIU) footbridge if they properly followed the civil engineers' plans.[29] Even though they built the material thing, those in the construction crew are not morally responsible for the collapse since they simply followed orders.[30] The engineers (more likely, the project managers) are morally responsible because they created faulty plans for the material builders to follow. This is the first case where someone is not morally responsible for what they build: if one builds by simply following instructions set by someone else. Most often, this occurs when a material builder is following the instructions of an immaterial builders, i.e. a designer or engineer.

There are situations where one cannot be blamed for their actions. Involuntary actions are not blameworthy. For Aristotle, "things [actions] coming about by force or because of ignorance are involuntary" (Aristotle 2000, p.30). Regarding forces, this is restricted to physical forces. For example, if a man is walking and "a wind or people who have [him] in their control were to carry him off," his movements would be involuntary (p.30). Regarding ignorance, Aristotle allows that "action done *in* ignorance" may be voluntary, like actions done while drunk (p.32, emphasis mine). These actions are still voluntary if one chose to get drunk. Only action done *by* ignorance, i.e. lapse of knowledge, renders an action involuntary. Furthermore, Aristotle writes that this is not "ignorance of the universal;" rather, one

[29] http://abcnews.go.com/US/pedestrian-bridge-florida-international-university-collapses/story?id=53774444

[30] I am not sure if this was actually the case. I am only speculating and hypothesizing about if this were the case.

must be ignorant "of the particulars which the consists in" (p.32).[31] For example, Aristotle writes that if you stab someone with a foil that you thought had a dull point at the end, this action is involuntary because you did not know it was sharp.

As another example, Aristotle writes that if you accidentally discharge a cannon when you intended to simply demonstrate how it works, this action is involuntary. If the cannonball hit and killed someone, Aristotle would say the person who accidentally discharged the cannon was not blameworthy for their actions because the discharge was involuntary out of ignorance. Aristotle seems too lenient towards accidents. Accidents may be actions done out of ignorance or they may be done unintentionally (often times, they are both). Regarding intentions, many would say that actions done intentionally may be blameworthy. If they are unintentional, there may still be cases where one may be culpable for their actions. Michael Bratman (1987) considers two types of actions where the outcomes may be expected: ones where the outcome is expected and intention (such as a terrorist bombing a school) or ones where the outcome is expected, but unintentional (such as a government bombing a group of enemies so near a school that they know that the school will be destroyed, but it is not their target) (p.139). I think both types of actions are blameworthy. Thus, all actions done unintentionally where the outcome is expected may be blameworthy (all other things permitting culpability).

Regarding ignorance, I think there is a reasonable expectation that the soldier that discharges a cannon be more careful around the cannon so as not to discharge it. Even if he discharges it on accident, this seems like a form of negligence. Gideon Rosen (2002) writes that: "[w]hen a person acts from ignorance, he is culpable for his action only if he is culpable for the ignorance from which he acts" (p.61). Rosen gives an example of where he is culpable for his ignorance:

> I am under an obligation to look out for other people when I'm out walking. If I recklessly shirk that obligation and wind up ignorant as a result, *the ignorance itself is culpable*, and in that case it's no excuse (p.63).

[31] Ignorance of the universal means something like not knowing what the good is and acting on that. For example, if you (incorrectly) thought stealing was good and stole, this would be ignorance of the universal.

The reckless walker fails to act with an appropriate amount of care for other people; when he does not notice other people and runs into them, he acts out of negligence. He should be blamed for his actions. Joseph Raz (2010) calls this obligation to exercise an appropriate amount of care a "duty not to harm through carelessness," which is a duty not to violate a "duty of care" towards others (p.9). By violating these duties and acting carelessly, i.e. acting negligently, one may be culpable for their actions. Thus, in order to be culpable for one's ignorance, they must be ignorance due to negligence. In general, one may be blamed for actions done out of ignorance if and only if this ignorance is negligent.

Negligence is especially pertinent when considering building. It seems that the designers of the seatbelt built out of ignorance, since they did not think of pregnant women when they designer the seatbelt. However, they ought to have thought of this specific, sensitive audience to design for. Same with the designers of optical character recognition software that could not recognize black faces. Even though acted out of ignorance, the designers of these products did so negligently and their actions are still blameworthy. Presumably, the seatbelts were made as such because there were no (or at least very few) women on the design staff; there were few black faces in the datasets that optical character recognition algorithms train on and relatively few black people in software design in general. This is why achieving adequate representation of gender, race, class, languages, religion, etc. among engineers and designers is important: in order to limit the amount of negligent harm targeted at specific socio-economic groups of people via built products. A builder is blameworthy for what they build unless they are non-negligently ignorant of the consequences of the thing they build.

Let us return to the FIU footbridge. At the time of writing, the only public statement about the bridge was that it collapsed due to cracks in the support beams (Fagenson 16 March 2018). I will present two hypothetical scenarios to illustrate the difference between negligent and non-negligent ignorance in building. First, suppose that the civil engineers who designed the bridge did not know the average number of people that walked over the bridge on a daily basis. They underestimated the weight that the bridge holds and cracks developed. Eventually, the cracks caused the bridge to collapse. In this case, it is reasonable to assume that knowing the average amount of daily traffic is necessary for designing a proper

bridge. Lack of such knowledge is a form of negligence. The engineers were ignorant of the average weight that the bridge ought to hold. However, their ignorance arose out of negligence. Thus, they are still morally responsible for the deaths and injuries caused by the collapse.

Second, suppose that the engineers did all they could to research the proper information about the bridge: they conducted studies to find the expected average and maximum amount of people who crossed the bridge, the weather conditions, etc. However, a pack of squirrels decided to take refuge under the bridge. They burrowed into support beams and caused the cracks. The cracks compromised the structural integrity of the bridge and it collapsed. In this case, there is no way that the engineers could have expected this problem. Thus, they built the bridge by non-negligent ignorance of the way that the squirrels would cause the bridge collapse. In general, a builder is not morally responsible for what they build if they do so by non-negligent ignorance of particular consequences. Ignorance of particulars is non-negligent when the particulars cannot be reasonably expected.

Section Six: Moral Responsibility for Built Agents

In an article published in *Ethics and Information Technology*, Andreas Matthias (2004) makes a similar point about exceptions to ascribing responsibility to those who engineer machines and algorithms. Matthias agrees that those who build machines ought to be responsible for the actions that machines do or the actions that operators of machines –assuming that the operator follows the manufacturer's instructions correctly (p.2-3). Matthias also agrees with me that the designer of a machine ought not to be held responsible for any "unforeseeable development" (p.2). Matthias argue that control is necessary for responsibility. Thus, when an engineer is "not capable of predicting… future machine behaviour," the engineer, not in control of these behaviors, is not responsible (p.2).

Citing Fischer and Ravizza (1998), Matthias writes that someone is in control of their action "only if he knows the particular facts surrounding his action, and if he is able to freely form a decision to act, and to select one of a suitable set of available alternative actions based on these facts" (p.2). There are two problems with this view. First, this is a misreading of Fischer and Ravizza (1998), who do not see

alternative actions as necessary for the kind of control necessary for moral responsibility, i.e. guidance control. What Matthias describes here is regulative control, which Fischer and Ravizza argue is not necessary for moral responsibility. (See Section 1 of Chapter 2 for a longer discussion of Fischer and Ravizza). Second, Matthias falls into the same trap as Aristotle, wherein he describes the lack of information –ignorance-- without considering ignorance that arises from negligence. Thus, Matthias' point about excusing software engineers for unpredictable consequences aligns with my general claims that we are not responsible for what we build, as long as Matthias' notion of control is understood as guidance control and ignorance as non-negligent.

As I noted when introducing algorithmic ethics, algorithmic technologies are unique in that they can make decisions without a human operator. In Chapter Two, I defined an actor as *autonomous* if they can interact with and adapt their actions in response to the environment, where their actions are not entirely determined by a designer. Traditional algorithmic technologies are not autonomous, since their algorithms are explicitly-programmed instructions that guide their decisions. However, a new kind of algorithms, machine learning algorithms, have allowed for the development for autonomous artificial agents.

Whereas traditional algorithms behave based on rules defined by software engineers, machine learning algorithms define their own rules by "generalizing from examples" (Domingos 2012, p.1). This entails "learning" patterns and generalizations within large datasets in order to identify a proper decision procedure (p.1). In a forthcoming work, Diane Michelfelder and Logan Stapleton use the example of a postal service algorithm that reads handwritten postage and translates the handwriting to computer characters. It would be impossible to do this using a traditional algorithm: the amount of variety and complexity in our handwriting is too difficult to put into explicit rules. Instead, we use learning algorithms which can simulate an artificial learning process and eventually define its own rules for deciding which handwritten character is which.

The learning algorithm is given a large dataset that contains, say, 10,000 pixelated snapshots of handwritten letters and numbers (for example, one of NIST's Special Databases). Each of these snapshots

may be labeled with the correct character or left unlabeled. If the snapshots are labeled, the learning algorithm parses through the dataset and finds patterns between all similarly-labeled snapshots, e.g. all the pixelated snapshots that correspond to the letter 'G'. If the snapshots are unlabeled, the learning algorithm parses through the dataset, picking out similarities between certain groups of pixels and grouping snapshots together based on these similarities. For example, a learning algorithm might put the letter 'I' and the number '1' in the same group, since many people write these two characters very similarly. In general, a learning algorithm takes in a large dataset and parses through it to pick out patterns and define its own decision procedures.

Two things become apparent about learning algorithms, following this explanation. First, technologies that use machine learning are capable of being autonomous. Luciano Floridi (2013) notes that the level of autonomy of an artificial agent seems to be proportional to the level of opacity of the algorithmic system that the agent runs on. An algorithm is *opaque* if a user cannot understand how the algorithm decides. The more opaque an algorithmic system, the less we understand its actions, and the more autonomous the artificial agent seems to be. Frank Pasquale (2015) details a number of ways an algorithm may be opaque: due to corporate secrecy, technical illiteracy, etc. In all of these forms of opacity for traditional algorithms, there is a way to demand transparency and be able to understand explicitly how an algorithm makes decisions.

Machine learning algorithms, however, are used precisely in situations where traditional algorithms cannot, because engineers cannot "provide an explicit, fine-detailed specification of how [complex tasks] should be executed" (Shalev-Shwartz & Ben-David, 2014 p.viii). Because an explicit specification of a task cannot be supplied, explicit explanations for the behavior of machine learning algorithms also cannot be supplied. Users cannot exactly know "the rationale of decision making rules produced by the algorithm" (Mittelstadt et al. 2016, p.3). Jenna Burrell (2016) argues that this leads to a new form of opacity where a learning algorithm automatically defines its own way to make decisions in a way that cannot be directly explained not only by a user, but even by its own designer. No outside observer can say exactly why an algorithm decides in a certain way or another; they can only give implicit

explanations based on the programming of the learning algorithm and its past behaviors. Technologies that run on learning algorithms can respond to their environment, in that they take in contextual data and change their decision procedures according to it. Yet, they can act independently of their environment and their designer: their designer cannot even directly explain why their algorithm acts the way it does. Under my definition of autonomy, learning algorithms allow for a technology to act autonomously.

Second, training a machine learning algorithm is not like constructing a basic machine or an explicitly-programmed algorithm. Though it is not apparently material form of building, creating software based on traditional algorithms is more like construction than it is like cultivation: nothing grows when you program a technology. For example, if you build a robot A that is programmed explicitly to tweet once every hour on the hour, you have built essentially the same thing as a cuckoo clock. Cuckoo clocks are constructed; as are robots programmed explicitly to behave like a cuckoo clock, like robot A does.

Suppose, for contrast, that you build a robot B that takes in a large dataset of pictures of the sun at various points throughout the day, labeled with their corresponding times, and learns to identify the time simply from looking at the sun in the sky. You can use robot B for the same purpose as robot A or the cuckoo clock; robot B will be able to tweet once every hour on the hour, by looking at the sun. Robot B is built in a different way than robot A. I think the ways that these two robots are built are different in kind. Robot A is constructed. The designer of robot B may give the robot a form by steering it towards a behavior of being able to identify the time given images of the sun. Yet, the designer of robot B did not explicitly construct a way to get there. The designer of robot B lets the algorithm loose in a large dataset to train itself the proper decision procedures. Training a machine learning algorithm is more like baking bread than it is like tossing a salad. The salad is constructed, since one simply puts a given number of ingredients in a bowl together and nothing grows. The bread is formed on the backs of the rising yeast. The baker does not *make* the bread in that the bread really comes into existence while the yeast eats the flour. Yet, the baker can be said to make the bread in that she gives a form to the yeast to grow into. Similarly, the learning algorithm is formed on the backs of the patterns in the dataset. Building a machine learning algorithm is a form of cultivation. Matthias goes so far as to call technologies built on learning

algorithms “software organisms” (p.15).

Most of modern robotics is based on machine learning. Thus, modern robotics ought to be thought of as a form of cultivation. We build a robot not by constructing it into the exact structure that we intend to. We build a robot by coaxing it into a form that we intend. In this sense, a robot can be said to grow. And a engineer can be said to raise a robot. The engineer cultivates an autonomous agent. This is like child-rearing. Child-rearing is a form of building, of shaping a living person into the form that they will become, of cultivating a person.

As such, the parent as a builder may be morally responsible for the thing that they build, i.e. the child. Intuitively, this is how we understood blameworthiness for children.

There are instances where one is not blameworthy for the actions of their child, however. Clearly, the parent is not responsible for the actions of their child once their child becomes a fully rational adult. As a rational adult (say, an adult capable of guidance control and second-order volition,) the child is a moral agent. Thus, the child may be blamed for their actions. No longer can the parent be blamed for their child’s actions. The blameworthiness of a parent diminishes once their child becomes a moral agent.

This is true of robotics as well. Deborah G. Johnson (2006) writes that “attributing independent moral agency to computers is dangerous because it disconnects computer behavior from human behavior” (p.204). She argues that attributing moral agency to computer systems redirects the ethical attention away from humans and onto computers (p.204). Amanda Sharkey (2017) argues that we ought not build artificial moral agents because this would “offload blame for mistakes or bad consequences onto robots” (p.210). Many robot ethicists argue against creating more sophisticated robots in the future for the fear that the human designers’ responsibility will be diminished once an artificial agent may be considered a moral agent. At the point where we may build moral agents, the moral responsibility for what we build is diminished like a parent whose child grows into a moral agent. Thus, it seems that there are no pressing worries about diminishing responsibility of robotics engineers until the far-off possibility that robots become moral agents. Many robot ethicists argue that we ought to prevent the creation of such robots (see the work of Noel Sharkey and the ‘Campaign for Killer Robots’).

I disagree with this. A designer's responsibility for the artificial agents they build lies long before the creation of moral agents. For example, the (1978) movie *Halloween* opens on a scene where the six-year-old Michael Myers (who goes on to be the silent serial killer in the movie) horrifically kills his sister with a kitchen knife. An atrocity is committed by a child. Who is to blame?

Michael Myers is not a full moral agent. It is reasonable to say that Michael does not understand his own actions. Thus, he cannot be blamed. His parents have raised him. Under the usual formulation of moral responsibility for a child's actions, Michael's parents seem to be blameworthy for Michael's actions. Yet, intuitively this seems false. Michael's parents had no way of knowing or any way of predicting that Michael would do something like this. This was an unforeseeable development (in Matthias' words) in the actions of the agent that they are cultivating. Thus, for the same reasons that the civil engineers were not responsible for the collapse of the Florida footbridge in my hypothetical scenario where squirrels caused cracks, i.e. because the engineers built out of non-negligent ignorance, we cannot blame Michael's parents for the actions of their child. Since neither Michael nor his parents are blameworthy, no one is to blame for the death of Michael's sister. Someone may be responsible for Michael Myers in a forward-looking sense only. In the movie, Michael is sent to an asylum to be rehabilitated or confined in order to prevent future deaths. (Spoiler: he escapes and starts terrorizing a town on Halloween night.) However, no one is blameworthy.

This is a case where those who build (Michael's parents) an agent (Michael) are entirely diminished of the moral responsibility for their agent. Yet, this is not a case like those outlined above, where responsibility is diminished because the agent becomes a full moral agent. Michael is not a moral agent. Yet, Michael's parents responsibility is still entirely diminished. We can imagine similar events for the builders of autonomous artificial agents. The artificial agent may not be considered a moral agent; thus, we cannot blame it. However, there may be situations where the builder may not be blameworthy as well. If a builder builds an artificial agent under non-negligent ignorance of unforeseen future behaviors of this agent, then the builder may not be responsible. Since the artificial agent is not a moral agent, it may not be morally responsible as well. Thus, there are situations where no one can be blamed, even

though the built agent is not a full moral agent.

Matthias (2004) argues for the same results. He argues that learning algorithms will lead to the development of machines whose actions lie outside of the control of their builders (p.3). He gives an example where NASA loses most of their ability to communicate with a semi-autonomous Mars rover; the rover, driving based on its own autonomous algorithmic control system, drives into a hole (p.3). In this case, the engineers were "not capable of predicting… future machine behaviour" (p.2). Thus, the engineer is not responsible. In fact, no one is to blame (p.3). Matthias (2004) calls this the *responsibility gap* (p.4).

The responsibility gap that Matthias and myself are arguing for is particularly striking. The problem with focusing on the day that artificial agents may be considered full moral agents or the kind of robot apocalypse that popular culture fixates on (see the 1984 movie, *The Terminator,* or Elon Musk's recent comments about AI) is that it diverts attention away from the ethical problems that arise with technologies that are possible today or in the near future.[32] It is a tragedy that Michael Myers can kill his sister and no one is to blame. But, it is harrowing to imagine a scenario where someone may be killed by a machine --designed and built by someone-- and no one is to blame.

Aside from their arguments about diminishing responsibility, both Johnson and Sharkey argue against attributing moral agency to robots because blaming human designers leads to forward-looking responsibility. Johnson argues that when we focus moral attention solely on the humans that make computer systems, "the design of computer systems is more likely to come into the sights of moral scrutiny, and, most importantly, better designs are likely to be created" (p.204).

However, I think blaming the designers of autonomous artificial agents is not the only way to effect forward-looking moral responsibility. Imagine a robot that is a full moral agent, able to control and understand the moral weight of their actions, that kills a human being. This scenario is scary. But, their programming can be changed and they may learn to act better. At the very least, we may blame the robot.

[32] www.cnbc.com/2018/04/06/elon-musk-warns-ai-could-create-immortal-dictator-in-documentary.html

Now imagine a robot that is not a moral agent that kills someone. For example, we might consider a scenario like the recent fatal accident that killed the driver of an autonomous Tesla (Tesla Team 27 March 2018). It is unclear whether the Tesla designers acted out of non-negligent ignorance of a scenario like this. However, if we assume that they did, then the designers could would be morally responsible for the driver's death. A human being's life ends. It is difficult to see why this accident happened. And no one is to blame. This second scenario seems scarier to me than the first. Considering the Tesla example, it is also clear that scenarios like this are already here and will increase in the future. These scenarios are bad enough to inspire forward-looking responsibility for robot engineers in order to avoid it. Thus, even though blame may not be placed on a designer, we may place forward-looking responsibility on the designers of such machines in order to prevent horrible scenarios where no one is to blame.

Conclusion

In this chapter, I defined *building* as action which brings something into being. This can be understood as construction, which makes something out of things that do not grow, or cultivation, which cares for and gives form to something that grows. Based on this definition, I presented a theory of why we are morally responsible for what we build. This kind of moral responsibility, I argue, is different in kind than individual, collective, or distributed responsibility. As well, there are scenarios where we are not responsible because of the same reasons that an actor is not blameworthy for involuntary actions. A builder is not responsible for a consequence or action of the thing they build if the builder cannot reasonably foresee this consequence or action. In this case, the builder builds out of non-negligent ignorance and is not blameworthy. With our most advanced technological capabilities, such as sophisticated learning algorithms, I argued that engineers are capable of building autonomous artificial agents (which are not yet moral agents). In general, engineers are responsible for the actions of the agent they build. I argued that there are certain cases where a builder builds an agent that commits an otherwise morally reprehensible act for which no one is to blame, since the artificial agent is not a full moral agent

and the builder builds the artificial agent out of non-negligent ignorance.

These cases are scary. They are imminent. As Matthias (2004) points out, these scenarios pose a "threat to both the consistency of the moral framework of society and the foundation of the liability concept in law" (p.4). More importantly, their possibility means that there will be scenarios where atrocities are committed by artificial agents and no one is to blame. We can imagine this happening today with the advent of sophisticated driverless cars. Thus, in order to prevent such scenarios of blameless atrocities, I think that engineers ought to adopt similar virtues that Vogel proposes as environmental virtues: self-knowledge and humility. Engineers ought to know themselves and their designs so as to understand where atrocious accidents may occur. This is to prevent negligence. However, they also ought to understand that horrible accidents do happen and it is better to be safe than sorry. Thus they ought to be humble and not design technologies so quickly that they lose the ability to predict certain horrific accidents.

Works Cited

1. Aristotle, trans. Irwin, Terence. 2000. *Nicomachean Ethics (Second Edition)*. Indianapolis: Hackett Publishing.
2. Asaro, P. M. 2006. “What should we want from a robot ethic?” *IRIE International Review of Information Ethics*, *6* (12):9-16.
3. Bekoff, Marc and Pierce, Jessica. 2009. *Wild Justice: The Moral Lives of Animals*. Chicago: University of Chicago Press.
4. Bendel, Oliver. 2016. “Considerations about the relationship between animal and machine ethics.” *AI & Society* 31 (1):103-108.
5. Bendik-Keymer, Jeremy. 2016. “Thinking Like a Mall: Environment Philosophy after the End of Nature [Book Review].” *Environmental Ethics* 38 (4): 507-508.
6. Bratman, Michael. 1987. *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
7. Burrell, Jenna. 2016. “How the machine ‘thinks’: Understanding opacity in machine learning algorithms.” *Big Data & Society* 3 (1): 1-12. doi: 10.1177/2053951715622512.
8. Carrington, Damian. 2016. “The Anthropocene epoch: scientists declare dawn of human-influenced age.” *The Guardian* (Aug 29). https://www.theguardian.com/environment/2016/aug/29/declare-anthropocene-epoch-experts-urge-geological-congress-human-impact-earth.
9. Chakrabarty, Dipesh. 2000. *Provincializing Europe: Postcolonial Thought and Historical Difference*. Princeton: Princeton University Press.
10. Church, Russell. 1959. “Emotional reactions of rats to the pain of others.” *Journal of Comparative and Physiological Psychology* 52 (2):132-134. doi:10.1037/h0043531.
11. Danaher, John. 2015. “The Philosophical Importance of Algorithms.” *Philosophical Disquisitions* (blog), July 20. philosophicaldisquisitions.blogspot.com/2015/07/the-philosophical-importance-of.html?m=1

12. de Waal, Frans. 2006. *Primates and Philosophers: How Morality Evolved*. Princeton: Princeton University Press.
13. DeGrazia, David. 2002. *Animal Rights: A Very Short Introduction*. Oxford University Press.
14. Dodig-Crnkovic, Gordana and Persson, Daniel. 2008. “Sharing Moral Responsibility with Robots: A Pragmatic Approach.” *Frontiers in Artificial Intelligence and Applications Volume 173* edited by Anders Holst, Per Kreuger, and Peter Funk, 165-168. Netherlands: IOS Press Amsterdam.
15. Domingos, Pedro. 2012. “A few useful things to know about machine learning.” *Communications of the ACM* 55 (10): 78-87. doi: 10.1145/2347736.2347755.
16. Dretske, Fred. 1993. “Can Intelligence Be Artificial?” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 71 (2):201-216.
17. Dyer-Witheford, Nick and de Peuter, Greig. 2009. *Games of Empire: Global Capitalism and Video Games*. Minneapolis: University of Minnesota Press.
18. Dyer-Witheford, Nick. 2015. *Cyber-Proletariat: Global Labour in the Digital Vortex*. London: Pluto Press.
19. Elden, Stuart. 2006. “Heidegger’s Animals.” *Continental Philosophy Review* 39 (3): 273-291. doi:10.1007/s11007-006-9020-7.
20. Elliot, Robert. 1982. “Faking Nature”. *Inquiry* 25 (1): 81-93. doi:10.1080/00201748208601955.
21. Fagenson, Zachary. 2018. “Engineer reported cracks in bridge before fatal collapse: transportation agency.” *Reuters*, March 16. www.reuters.com/article/us-florida-bridge/engineer-reported-cracks-in-bridge-before-fatal-collapse-transportation-agency-idUSKCN1GS16M.
22. Fischer, John M. and Ravizza, Mark. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

23. Floridi, Luciano. 2013. “The morality of artificial agents.” In *The Ethics of Information*, edited by Luciano Floridi, 134-159. Oxford: Oxford University Press.
24. Floridi, Luciano and Sanders, J. W. 2004. “On the morality of artificial agents.” *Minds and Machines* 14 (3):349-379.
25. Frankfurt, Harry. 1971. “Freedom of the will and the concept of a person.” *Journal of Philosophy*. 68 (1): 5-20.doi:10.2307/2024717.
26. Goodman, Bryce W. 2016. “A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection.” In *Neural Information Processing Systems Symposium Proceedings: ML and the Law, Barcelona, 2016.*
27. Gottfried, Mara H. May 24 2016. “St. Paul mom had just taken son to bus before being hit, police say.” *Pioneer Press*. www.twincities.com/2016/05/24/st-paul-driver-who-hit-mom-in-crosswalk-thought-stopped-car-was-turning-police-say/.
28. *Halloween*. 1978. Directed by John Carpenter. Kansas City: Universal Pictures.
29. Han, Byung-Chul, trans. Butler, Erik. 2017. *In the Swarm: Digital Prospects.* Cambridge: The MIT Press.
30. Hardt, Michael. 2005. “Immaterial Labor and Artistic Production, Rethinking Marxism, 17 (2):175-177. doi:10.1080/08935690500046637.
31. Hardt, Michael and Negri, Antonio. 2000. *Empire*. Harvard University Press.
32. Heflin, Ashley Shew. 2011. *A Unifying Account of Technological Knowledge: Animal Construction, Tool Use, and Technology*. Ph.D. Virginia Polytechnic Institute and State University.
33. Heidegger, Martin, trans. Gray, J. Glenn. [1968] 1976. *What is Called Thinking?* New York: Perennial Press.
34. ____, trans. Stambaugh, John. [1953] 2010. *Being and Time*. New York: State University of New York Press, Albany.

35. ____, trans. Hofstadter, Albert. [1971] 2013. “Building Dwelling Thinking.” In *Poetry, Language, Thought*, edited by Albert Hofstadter. 141-159. New York: Harper Collins.
36. ____, trans. McNeill, William and Walker, Nicholas. 1995. *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude*. Indiana: Indiana University Press.
37. Hellström, Thomas. 2013. *Ethics and Information Technology* 15 (2):99-106. doi:10.1007/s10676-012-9301-2.
38. Hill, Robin. 2016. “What an Algorithm Is.” *Philosophy & Technology* 29 (1): 35–59.
39. Hoły-Łuczaj, Magdalena. Forthcoming. “In Search of Allies for Postnatural Environmentalism, or Revisiting an Ecophilosophical Reading of Heidegger,” in *Environmental Values*. White Horse Press.
40. Johnson, Deborah G. 2006. “Computer Systems: Moral entities but not moral agents.” *Ethics and Information Technology*. 8:196-204. doi:10.1007/s10676-006-9111-5.
41. Katz, Eric. 1992. “The Big Lie: Human Restoration of Nature.” *Research in Philosophy and Technology* 12: 231-241; reprinted in William Throop (ed.). 2000. *Environmental Restoration: Ethics, Theory and Practice*. New York: Humanity Books. 83–93.
42. ____. 2010. “The Nazi Engineers: Reflections on Technological Ethics in Hell.” *Science & Engineering Ethics* 17:571-582. doi:10.1007/s11948-010-9229-z.
43. Korsgaard, Christine M. 1996. *The Sources of Normativity*. Massachusetts: Cambridge University Press.
44. Kraemer, Felicitas, van Overveld, Kees, and Peterson, Martin. 2010. “Is there an ethics of algorithms?” *Ethics and Information Technology* 13 (3): 251-260.
45. Latour, Bruno. 1999. *Pandora’s Hope: Essays on the Reality of Science Studies. Cambridge*. Massachusetts: Harvard University Press.
46. Lazzarato, Maurizio. 1996. “Immaterial Labor,” in *Radical Thought in Italy: A Potential Politics*, edited by Michael Hardt and Paolo Virno, 133-150. Minneapolis: University of Minnesota Press.

47. Leopold, Aldo. 1949. *A Sand County Almanac, and Sketches Here and There*. New York: Oxford University Press.
48. Marx, Karl. [1844] 1975. “Economic and Philosophical Manuscripts (1844)” in Marx, Karl; Benton, Gregor; and Livingstone, Rodney. *Early Writings [of] Karl Marx*. 1975. Harmondsworth: Penguin: 279-400.
49. ____. [1844] 1972. “On the Jewish Question” in Marx, Karl; Engels, Friedrich; and Robert C. Tucker. 1972. *The Marx-Engels Reader*. New York: Norton.
50. ____. [1858]. *Grundrisse*. www.marxists.org/archive/marx/works/1857/grundrisse/. Accessed Apr 2018.
51. ____. [1867]. *Capital, Volume One*. www.marxists.org/archive/marx/works/1867-c1/ch07.htm. Accessed Apr 2018.
52. Matthias, Andreas. 2004. “The responsibility gap: Ascribing responsibility for the actions of learning automata.” *Ethics and Information Technology* 6 (3):175-183.
53. McKibben, Bill. 1989. *The End of Nature*. New York: Anchor Books.
54. Mittelstadt, Brent D., Allo, Patrick, Taddeo, Mariarosaria, Wachter, Sandra, and Floridi, Luciano. 2016. “The Ethics of Algorithms: Mapping the Debate.” *Big Data & Society* 3 (2): 1-21. doi:10.1177/2053951716679679.
55. Morton, Timothy. 2013. *Hyperobjects: Philosophy and Ecology after the End of the World*. Minneapolis: University of Minnesota Press.
56. Musschenga, Bert. 2015. “Moral Animals and Moral Responsibility.” *Les ateliers de l'éthique/The Ethics Forum* 10 (2):38-59.
57. Nihlén Fahlquist, Jessica. 2017. “Responsibility Analysis.” In *The Ethics of Technology: Methods and Approaches*, edited by Sven Ove Hansson, 129-142. London: Rowland & Littlefield.
58. Nissenbaum, Helen. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto: Stanford University Press.

59. Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engineers Reinforce Racism*. New York: New York University Press.
60. O'Neill, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.
61. Pariser, Eli. 2011. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. London: Penguin Press.
62. Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.
63. Pluhar, Evelyn. 1995. *Beyond Prejudice: The Moral Significance of Human and Nonhuman Animals*. Duke University Press.
64. Raz, Joseph. 2010. "Responsibility and the Negligence Standard." *Oxford Journal of Legal Studies*. 30 (1):1-18. doi:10.1093/ojls/gqq002.
65. Regan, Tom. 1983. *The Case for Animal Rights*. University of California Press.
66. Rice, George E. and Gainer, Priscilla. 1962. "'Altruism' in the albino rat." *Journal of Comparative and Physiological Psychology* 55(1):123-125. doi:10.1037/h0042276.
67. Rosen, Gideon. 2002. "Culpability and Ignorance." *Proceedings of the Meeting of the Aristotelian Society*. University of London.
68. Rosenberger, Robert. 2017. "Phenomenological Approaches to Technological Ethics." In *The Ethics of Technology: Methods and Approaches*, edited by Sven Ove Hansson, 67-82. London: Rowland & Littlefield.
69. Rowlands, Mark. 2012. *Can Animals Be Moral?*. New York: Oxford University Press.
70. Shalev-Shwartz, Shai and Ben-David, Shai. 2014. *Understanding Machine Learning: From Theory to Algorithms*. New York: Cambridge University Press.
71. Shapiro, Paul. 2006. "Moral agency in other animals." *Theoretical Medicine and Bioethics* 27 (4):357-373.

72. Sharkey, Amanda. 2017. “Can robots be responsible moral agents? And why should we care?” *Connection Science* 29(3):210-216. doi:10.1080/09540091.2017.1313815.
73. Shaver, Katherine. 2012. “Female dummy makes her mark on male-dominated crash test.” *The Washington Post*, March 25. www.washingtonpost.com/local/trafficandcommuting/female-dummy-makes-her-mark-on-male-dominated-crash-tests/2012/03/07/gIQANBLjaS_story.html?utm_term=.690d81b215e9.
74. Shew, Ashley. 2017. *Animal Constructions and Technological Knowledge*. London: Lexington Books.
75. Steffen, Will, Crutzen, Paul J., and McNeill, John R. “The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature?” *AMBIO: A Journal of the Human Environment* 36(8):614-621.
76. Sinnott-Armstrong, Walter. 2005. “It’s Not My Fault: Global Warming and Individual Moral Obligations,” in *Perspectives on Climate Change: Science, Economics, Politics, Ethics*, edited by Walter Sinnott-Armstrong and Richard B. Howarth, 293-315. Burlington, VT: Emerald Group Publishing.
77. Sullins, John P. 2006. “When is a robot a moral agent.” *International Review of Information Ethics* 6 (12):23-30.
78. Sunstein, Cass R. 2006. *Infotopia: How Many Minds Produce Knowledge*. New York: Oxford University Press.
79. *The Terminator*. 1984. Directed by James Cameron. Orion Pictures.
80. Tesla Team. 2018. “What We Know About Last Week’s Accident.” *Tesla* (blog), March 27. www.tesla.com/blog/what-we-know-about-last-weeks-accident
81. Vogel, Steven. 2015. *Thinking Like a Mall: Environmental Philosophy After the End of Nature*. MIT Press.

82. Wallach, Wendell and Allen, Colin. 2009. *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press.

83. Warren, Mary Anne. 1987. “Difficulties with the Strong Rights Position.” *Between the Species* 2(4):433-441.

84. Waters, Colin N., Jan Zalasiewicz, Colin Summerhayes, Anthony D. Barnosky, Clément Poirier, Agnieszka Gałuszka, Alejandro Cearreta, Matt Edgeworth, Erle C. Ellis, Michael Ellis, Catherine Jeandel, Reinhold Leinfelder, J. R. McNeill, Daniel deB. Richter, Will Steffen, James Syvitski, Davor Vidas, Michael Wagreich, Mark Williams, An Zhisheng, Jacques Grinevald, Eric Odada, Naomi Oreskes, Alexander P. Wolfe (2016). “The Anthropocene is functionally and stratigraphically distinct from the Holocene.” *Science* 351:6269. doi:10.1126/science.aad2622.