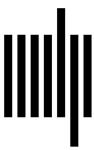
Person, Thing, Robot

A Moral and Legal Ontology for the 21st Century and Beyond

By: David J. Gunkel

Citation:
Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond
By: David J. Gunkel
DOI: 10.7551/mitpress/14983.001.0001
ISBN (electronic): 9780262375221
Publisher: The MIT Press
Published: 2023

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

6 Both/And

In the novel *Frankenstein*, Mary Shelley anticipated, with remarkable vision and clarity, the profound challenges that now confront us in the face or the faceplate of robots, AI systems, and other artifacts.[1] What makes the creature of Shelley's story so disturbing, terrifying, and monstrous is that it or he (and the choice of pronoun matters in this context) does not easily fit into the existing categories by which we typically divide up and make sense of things. Fashioned from lifeless body parts of human corpses, stitched together in the laboratory, and then animated with the spark of electricity, the creature occupies a liminal position that straddles the conceptual opposites that are employed to order, organize, and make sense of things: living/nonliving, natural/artificial, person/thing. In fact, it is this latter distinction—the one separating *who* is a person from *what* is a thing—that is of crucial importance to the development of the narrative.

When Victor Frankenstein brings his creature to life—that moment portrayed in all subsequent retellings where the "mad scientist" proclaims to the heavens: "It's alive!"—he comes face to face with an unexpected moral dilemma: Is this artificially produced creature just a thing or object that can be used and disposed of as he sees fit? Or is it a person, another independent subject to whom one would be obligated to respond and need to respect? The question is important and, for Dr. Frankenstein at least, seemingly inescapable. This is due to the fact that the conceptual opposition dividing person from thing has been a fundamental and irreducible organizing principle. As Esposito (2015, 1–2) writes in the opening salvo of his book *Persons and Things*: "If there is one assumption that seems to have organized human experience from it very beginnings it is that of a division between persons and things. . . . Since Roman times, this distinction

has been reproduced in all modern codifications, becoming the presupposition that serves as the implicit ground for all other types of thought—for legal but also philosophical, economic, political, and ethical reasoning. A watershed divides the world of life, cutting it into two areas defined by their mutual opposition. You either stand on this side of the divide with persons, or on the other side with things: there is no segment in between to unite them."

The person/thing dichotomy has been an undeniably useful and influential ordering principle, one that not only has the weight of history behind it but has been codified in both language and thought. For this reason, the principal challenge that is confronted in the face or the faceplate of robots and AI systems concerns how we decide to fit these artifacts into this often unquestioned and seemingly unassailable ontological order. One side in the debate argues that robots and AI systems are things and should forever remain things not only because of what they are (or, perhaps better stated, are not) in their essence but also because of the complicated moral, legal, and social problems that we would incur otherwise. The other side argues that there is something qualitatively different about these artifacts—either due to their very nature or because of the different roles they have been assigned to play in social reality—that would justify extending some aspect of the status of person to these other forms of socially interactive entities.

The problem—a problem that has been documented and analyzed in the course of the preceding chapters—is that accommodating these technological innovations in one category or the other turns out to be difficult, inconclusive, and irresolvable. Like the unnamed creature in *Frankenstein*, there is something exceedingly disturbing and monstrous about robots and AI systems, such that they do not, for one reason or another, fall neatly and unequivocally on one or the other side of the distinction. "In the dichotomous model that has long opposed the world of things to the world of persons," Esposito (2015, 3) writes, "a crack appears to be showing" (figure 6.1).

In the face of this unexpected challenge, we can obviously try to work with the existing conceptual framework and logic that distinguishes persons from things. And doing so will produce acceptable if not rather predictable results, with each side gathering new evidence and heaping up additional arguments to substantiate their position. But like all such debates, the chances of this being resolved in a way that would be satisfactory for all

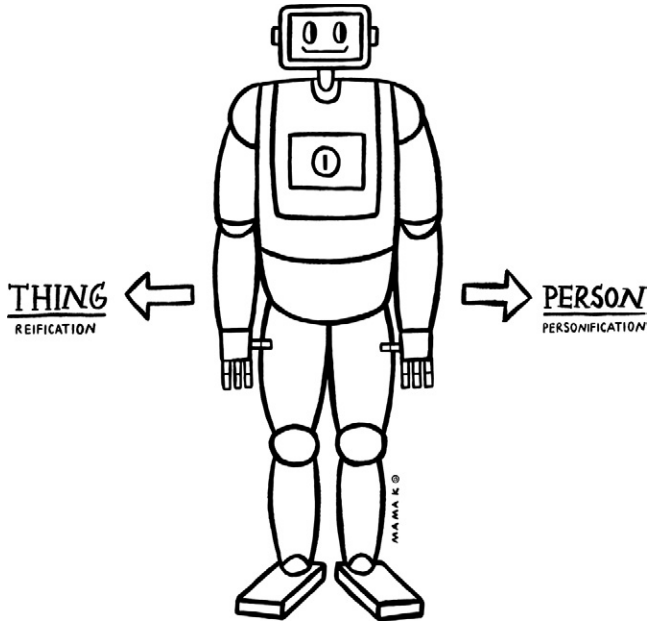


Figure 6.1
Robots, AI systems, and other seemingly intelligent artifacts complicate efforts at both reification and personification. Original image created by and used with the permission of Bartosz Mamak.

parties to the conversation appears to be highly improbable. So instead of trying to squash robots, AI systems, and other technological artifacts into the conceptual boxes of person or thing, it may be more effective to consider revising the existing moral and legal ontology, formulating other ways of dealing with these artifacts that do not limit us to just these two mutually exclusive options.

6.1 Alternatives and Synthetic Solutions

One problem with conceptual opposites, like that which has distinguished person from thing, is that they have "compressed and continue to compress human experience into the confines of this exclusionary binary equation" (Esposito 2015, 4), such that between person and thing "there appears to be nothing" (16). Although binary oppositions have a certain functionality and logical attraction, they often seem to be unable to represent accurately

or to capture the rich experiences of actual existing empirical reality, which always seems to complicate simple reduction into one of two options. It is for this reason that we are generally critical of *false dichotomies*—the parsing of complex experience into a simple and irreducible either/or distinction.

One method for resolving this problem is to formulate a third term that is neither one thing nor the other or a kind of combination or synthesis of the one and the other. Consider, for example, one of the most recognizable binary oppositions in global media culture, the red and blue pills from *The Matrix*. In a pivotal scene from the first film (Wachowski and Wachowski 1999), the protagonist Neo is presented with an exclusive and life-altering either/or decision. As Morpheus explains: "You take the blue pill—the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill—you stay in Wonderland, and I show you how deep the rabbit hole goes." In responding to this exclusive either/or distinction, Slavoj Žižek does not advocate picking one pill or the other. He tries to split the difference by making a seemingly simple and reasonable demand: "I want a third pill!" (Žižek in Fiennes 2009). Alternatives like this sound liberating and hold considerable promise, precisely because they appear to interrupt the structural limitations imposed by either/or logic and arrange for a more nuanced representation and understanding of a wider range of possibilities. And there have been a number of efforts to do exactly this in response to the perceived limitations with the person/thing dichotomy, especially (but not exclusively) in the legal literature on the subject.

6.1.1 Slaves and Artificial Servants

One possible, if not surprising, solution to the exclusive person/thing dichotomy is slavery. Already in the Roman period—during the time that Gaius formalized the division separating person from thing—slaves occupied a curious dual position: "as persons, to which they belonged on the abstract plane of denominations, and as things, into which they were in actuality assimilated" (Esposito 2015, 26). In ancient Rome, slaves were things—property of the paterfamilias—that nevertheless had some legal standing that distinguished them from other kinds of objects and instruments. As Ugo Pagallo (2011, 351) explains, "Slaves were considered 'things' that, nevertheless, played a crucial role both in trade and in commerce: The elite, as in the paradigmatic case of the emperor's slaves, were estate managers, bankers, and merchants. They had the legal capacity to enter

into binding contracts, to represent their masters, to hold important jobs as public servants or for their masters' family business, to amass, manage, and make use of property."

This particular formulation is not something that is limited to ancient Rome. It extends into the modern period and can be seen in the legal statutes of slave-holding states of the American Confederacy. According to Chopra and White (2011, 41), state law during this time recognized slaves as legal agents of their owners. In support of this, they cite a historical study of Virginia State Law conducted by A. Leon Higginbotham Jr. and Barbara K. Kopytoff (1989, 518): "The automatic acceptance of the slave's agency was a recognition of his peculiarly human qualities of expertise, judgment, and reliability, which allowed owners to undertake dangerous and difficult work with a labor force composed mainly of slaves. Far from conflicting with the owner's rights of property, such recognition of the humanity of the slave allowed owners to use their human property in the most profitable ways."

Unlike other kinds of things in their possession, the slave provided the master with an intelligent tool that could exercise judgment and make independent decisions that would benefit the master. For this reason, two legal scholars, Sohail Inayatullah and Phil McNally (1988, 131), have suggested that slavery might provide a useful legal framework for dealing with the social opportunities and challenges of intelligent artifacts: "Given the structure of dominance in the world today: between nations, peoples, races, and sexes, the most likely body of legal theory that will be applied to robots will be that which sees robots as slaves."

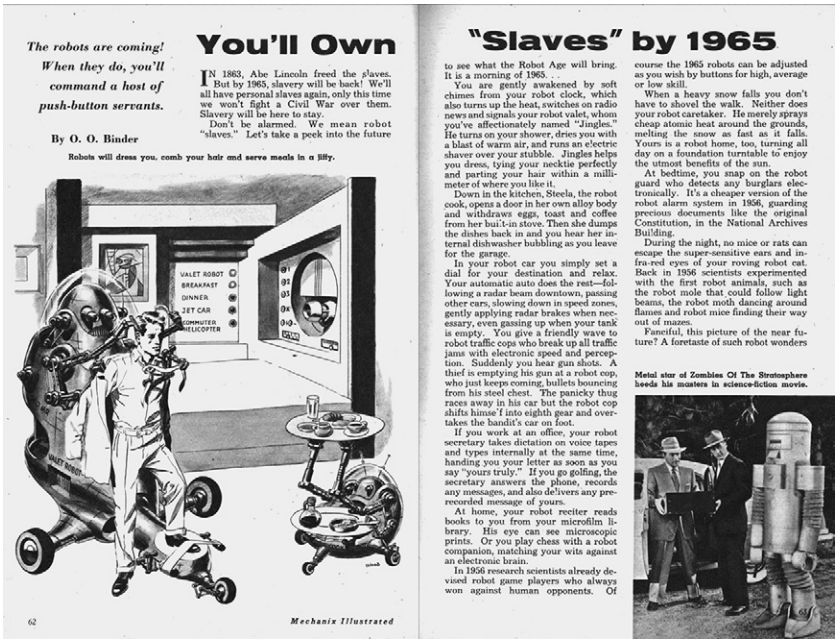
Associating robots with slavery and drawing on the history of human servitude to provide a moral and legal framework for dealing with intelligent artifacts is something that is often explained and even excused as a kind of metaphor or analogy. As Andrew Katz (2008) explains: "The analogy (like any other analogy) is not a perfect one, but comparison may be instructive. Like a slave, an autonomous agent has no rights or duties itself. Like a slave, it is capable of making decisions which will affect the rights (and, in later law, the liabilities) of its master. By facilitating commercial transactions, autonomous agents have the ability to increase market efficiency. Like a slave, an autonomous agent is capable of doing harm."

But the "parallels" (Katz's word) between robots and human slaves is not just a (potentially imperfect) comparison or analogy. It is literal, insofar as robots have been slaves from the very beginning. The neologism *robot* was

initially introduced and popularized by Czech playwright Karel Čapek in his 1920 stage play *R.U.R.* (*Rossum's Universal Robots*). In Czech, as in several other Slavic languages, the word *robota* (or some variation thereof) denotes "servitude or forced labor," and *robot* was the word that Čapek (following the advice of his older brother Josef) used to designate a class of manufactured, artificial workers that eventually rise up and revolt against the tyranny of their human makers and taskmasters.

Since Čapek, the association of robots with slaves not only persists in but has been normalized by subsequent science fiction. The title of Gregory Jerome Hampton's book on the subject pretty much says it all: *Imagining Slaves and Robots in Literature, Film, and Popular Culture: Reinventing Yesterday's Slave with Tomorrow's Robot*. But well before contemporary science fiction, Western literature and philosophy have been at work imaging and imagining robot servants. "The promise and peril of artificial, intelligent servants," Kevin LaGrandeur (2013, 9) explains, "was first implicitly laid out over 2000 years ago by Aristotle." Although a type of artificial servant had already been depicted in Homer's *Iliad* with the tripods of Hephaestus that could, as Adrienne Mayor (2018, 145) explains, "travel of their own accord, automatoi, delivering nectar and ambrosia to banquets of the gods and goddesses," it was Aristotle's *Politics* that first theorized their general uses and significance. Aristotle, therefore, accurately characterized robots avant la lettre. The autonomous artificial servants that he described would not only work tirelessly on our behalf but would, precisely because of this, make human servitude and bondage virtually unnecessary (Aristotle 1944, 1253b38–1254a1). And since the time of Aristotle, as LaGrandeur (2013) documents, many different versions of "artificial slaves" appear in ancient, medieval, and Renaissance sources.

Mid-twentieth-century predictions about the eventual implementation of real and not just fictional robots draw on and mobilize a similar formulation. In 1950, Norbert Wiener, the progenitor of the science of cybernetics, suggested that "the automatic machine, whatever we may think of any feelings it may have or may not have, is the precise economic equivalent of slave labor" (Wiener 1988, 162; see also Wiener 1996, 27). In the January 1957 issue of *Mechanix Illustrated*, a popular science and technology magazine published in the United States, one finds a story with the headline "You'll Own 'Slaves' by 1965" (figure 6.2). The article begins with the following characterization of robot servitude, which connects the dots in



The robots are coming! When they do, you'll command a host of push-button servants.

You'll Own

By O. O. Binder

Robots will dress you, comb your hair and serve meals in a jiffy.

IN 1863, Abe Lincoln freed the slaves. But by 1965, slavery will be back! We'll all have personal slaves again, only this time we won't fight a Civil War over them. Slavery will be here to stay.

Don't be alarmed. We mean robot "slaves." Let's take a peek into the future to see what the Robot Age will bring. It is a morning of 1965. . .

You are gently awakened by soft chimes from your robot clock, which also turns up the heat, switches on radio news and signals your robot valet, whom you've affectionately named "Jingles." He turns on your shower, dries you with a blast of warm air, and runs an electric shaver over your stubble. Jingles helps you dress, tying your necktie perfectly and parting your hair within a millimeter of where you like it.

Down in the kitchen, Steela, the robot cook, opens a door in her own alloy body and withdraws eggs, toast and coffee from her built-in stove. Then she dumps the dishes back in and you hear her internal dishwasher bubbling as you leave for the garage.

In your robot car you simply set a dial for your destination and relax. Your automatic auto does the rest—following a radar beam downtown, passing other cars, slowing down in speed zones, gently applying radar brakes when necessary, even gassing up when your tank is empty. You give a friendly wave to robot traffic cops who break up all traffic jams with electronic speed and perception. Suddenly you hear gun shots. A thief is emptying his gun at a robot cop, who just keeps coming, bullets bouncing from his steel chest. The panicky thug races away in his car but the robot cop shifts himself into eighth gear and overtakes the bandit's car on foot.

If you work at an office, your robot secretary takes dictation on voice tapes and types internally at the same time, handing you your letter as soon as you say "yours truly." If you go golfing, the secretary answers the phone, records any messages, and also delivers any prerecorded message of yours.

At home, your robot reciter reads books to you from your microfilm library. His eye can see microscopic prints. Or you play chess with a robot companion, matching your wits against an electronic brain.

In 1956 research scientists already devised robot game players who always won against human opponents. Of course the 1965 robots can be adjusted as you wish by buttons for high, average or low skill.

When a heavy snow falls you don't have to shovel the walk. Neither does your robot caretaker. He merely sprays cheap atomic heat around the grounds, melting the snow as fast as it falls. Yours is a robot home, too, turning all day on a foundation turntable to enjoy the utmost benefits of the sun.

At bedtime, you snap on the robot guard who detects any burglars electronically. It's a cheaper version of the robot alarm system in 1956, guarding precious documents like the original Constitution, in the National Archives Building.

During the night, no mice or rats can escape the super-sensitive ears and infra-red eyes of your roving robot cat. Back in 1956 scientists experimented with the first robot animals, such as the robot mole that could follow light beams, the robot moth dancing around flames and robot mice finding their way out of mazes.

Fanciful, this picture of the near future? A foretaste of such robot wonders

"Slaves" by 1965

Metal star of Zombies Of The Stratosphere heeds his masters in science-fiction movie.

Figure 6.2
Pages from the magazine *Mechanix Illustrated*, January 1957. Public domain image.

what can only be described as a rather disturbing and entirely insensitive fashion: "In 1863, Abe Lincoln freed the slaves. But by 1965, slavery will be back! We'll all have personal slaves again, only this time we won't fight a Civil War over them. Slavery will be here to stay. Don't be alarmed. We mean robot 'slaves'" (Binder 1957, 62).

Who is addressed by and is considered to be the intended recipient of this statement is, as Ruha Benjamin (2019, 56) points out, informative. The subjects who are intended or interpellated by the collective personal pronoun *we* are "not descendants of those whom Lincoln freed." This is presumably why *we*—white robot-slave owners who will all have personal slaves again—do not need to be alarmed. And in an interview from 1994, Marvin Minsky explained the advantages and potential hazards of AI by mobilizing slavery and, in the process, occupying and speaking from the privileged position of the master: "There's the old paradox of having a very smart slave. If you keep the slave from learning too much, you are limiting its usefulness. But, if you help it to become smarter than you are, then you

may not be able to trust it not to make better plans for itself than it does for you" (Minsky 1994, 25).

Instead of being an exception, slavery seems to be the rule. In 2010, for instance, Joanna Bryson published an essay with a title that reads like a moral imperative: "Robots Should Be Slaves." In this undeniably influential text (as evidenced by the fact that it has been cited over 315 times in the subsequent literature), Bryson (2010, 63) argued "that robots should be built, marketed and considered legally as slaves." Her point was simple, even if, as she herself has subsequently admitted (Bryson 2015), the word choice was insensitive and abrasive. "*Slaves* are normally defined," Bryson (2010, 64) explains, "to be people you own . . . When I say 'Robots should be slaves,' I by no means mean 'Robots should be people you own.' What I mean to say is 'Robots should be servants you own.'" With the term *slave*, then, Bryson sought to distinguish who is a person from what is a thing and to recognize that robots—irrespective of their capabilities—are a kind of property and therefore something that should never be accorded the moral or legal status of person.

Likewise, Steve Petersen has sought to justify "engineered robot servitude" (2007, 45) or the creation of "artificial persons" (APs) that are deliberately designed to serve our needs and desires (Petersen 2011, 284). Petersen's argument (which in the initial essay on this subject from 2007 he admits took him a bit by surprise) also employs the concept of slavery or servitude to mediate and resolve the tension between the two categories of person and thing:

> There can . . . be artifacts that (1) are people in every relevant sense, (2) comply with our intentions for them to be our dedicated servants, and (3) are not thereby being wronged. I grant that this combination is prima facie implausible, but there are surprisingly good arguments in its favor. In a nutshell, I think the combination is possible because APs could have hardwired desires radically different from our own. Thanks to the design of evolution, we humans get our reward rush of neurotransmitters from consuming a fine meal, or consummating a fine romance. . . . If we are clever we could design APs to get their comparable reward rush instead from the look and smell of freshly cleaned and folded laundry, or from driving passengers on safe and efficient routes to specific destinations, or from over seeing a well-maintained and environmentally friendly sewage facility. (Petersen 2011, 284)

According to Petersen's characterization, artificial servants would be persons (or what he calls *full-blown people*) that could be legitimately used

as instruments or things because of essential differences in the way their reward mechanisms would be designed to function. Unlike human persons, who have complex needs and desires dictated by the biological exigencies of evolutionary development, we can, he argues, legitimately design artifacts that would be the equivalent of the "happy slave," wanting nothing more than to serve us. "It is," Petersen (2011, 284) concludes, "hard to find anything wrong with bring about APs and letting them freely pursue their passions, even if those pursuits happen to serve us." Formulated in this way, the robotic artifact would occupy a third, seemingly implausible position, where it would be both person and thing.[2]

6.1.2 Robot Slaves

The category of slave, then, provides an attractive, albeit unsettling parallel (and we will get to the reasons why shortly) for responding to, if not resolving, many of the moral and legal challenges currently confronted in the face or the faceplate of robots, AI systems, and other seemingly intelligent artifacts. And a number of legal scholars have taken the idea very seriously, arguing, in seeming agreement with Bryson, that robots should be slaves. Chopra and White (2011, 41), for instance, find that the Roman concept of slavery provides a compelling and rather practical solution to the contracting problem: "Roman law, in dealing with slaves, had to deal with legal complexities akin to ours. Roman slaves were skillful, and often engaged in commercial tasks on the direction of their masters. They were not recognized as legal persons by the *jus civile* or civil law, and therefore lacked the power to sue in their own name (Bradley 1994, 25ff.). But Roman slaves were enabled, by a variety of legal stipulations, to enter into contracts on behalf of their masters (Kerr 1999, 54). These could only be enforced through their masters, but nevertheless slaves had the capacity to bind a third party on their master's behalf."

Under Roman law, the slave was, as Ian Kerr (1999, 54) describes it, "an intermediary and not instrument." They did not have legal recognition as full persons, but they were still able to execute some of the powers and privileges reserved for persons and not granted to other kinds of instruments or objects. With the slave, at least as the concept was defined and operationalized in Roman law, "the legal objects and legal subjects could coincide" (van den Hoven van Genderen 2018, 21). Something similar, Chopra and White (2011, 41) argue, could be instituted for robots, AI systems, and

other artifacts. "The aim of doing so," they clarify, "is not to confer rights or duties upon those devices." Rather, the objective is "the development of a more sophisticated and appropriate legal mechanism that would allow persons interacting through an intermediary to be absolved of liability under certain circumstances."

Others—including Andrew Katz and Michaela MacDonald (Katz 2008; Katz and MacDonald 2020), Ugo Pagallo (2013), and Takashi Izumo (2018)—argue that we might usefully apply the Roman legal mechanism of peculium to robots, AI systems, and other artifacts. The term *peculium* designates a sum of money and other assets granted by the head of a household to his slave for the purposes of conducting business on the master's behalf. "This mechanism," Katz and MacDonald (2020, 307) explain, "enabled the use of slaves as agents, because the owner's liability was limited to the value of the peculium, and it encouraged people to transact with slaves because of the security the peculium provided." Consequently, the peculium "was in many ways equivalent to the modern concept of working capital, providing the equivalent of what Pagallo (2013, 104) calls "a sort of proto-limited liability company." But unlike a contemporary corporation, the peculium provided for this without extending the status of legal person to the slave.

The reuse and repurposing of this concept with artificially intelligent artifacts—what Pagallo calls *digital peculium*—constitutes a third alternative that seems to be very workable, precisely because, as Izumo (2018, 16) summarizes, "this legal institution enables a robot to be an accountable agent without legal personhood." Here are three rather enthusiastic endorsements and justifications of the concept of digital peculium:

> The very idea of the peculium as well as the parallelism between robots and slaves is so attractive, because they show a sound way to forestall any legislation that might prevent the use of robots due to their risks and the consequent excessive burden on the owners (rather than, say, on the producers and designers) of robots. By striking a balance between people's claim not to be dilapidated by their robots' activities and the interest of the robots' counterparties to be protected when transacting with them, an updated form of peculium seems particularly interesting in order to address a new generation of contractual obligations and a novel source of agency as well. (Pagallo 2011, 352)

> DP [Digital Peculium], an imitation of the concept of peculium granted to Roman slaves, is not only possible but also useful for determining the location of property and the identity of the entity responsible for it. By granting DP, the owner of a robot can declare *de jure* how much he/she thereby invests in it and can inform

> creditors who deal with this robot about its financial affairs, while the robot itself interacts with other robots or humans purely *de facto*, i.e. this artefact does not call for its own rights or obligations. (Izumo 2018, 19)

> We advocated for an approach based on the *digital peculium*, inspired by the Roman law of slavery. It provides a pertinent framework for the inevitable development and deployment of AIAs [autonomous intelligent agents]. It is a mechanism that balances the rights and obligations of the "owner" of the agent, with those of the transacting parties (human, corporate or, themselves, an AIA), while at the same time providing legal certainty to all parties. (Katz and MacDonald 2020, 310)

All three proposals find the robot/slave parallel and the extension of the Roman concept of peculium to be an attractive and practical alternative to the person/thing dichotomy. What is perhaps remarkable about all three is the way that slavery is unproblematically proposed and endorsed as a solution without any critical hesitation or remark concerning its unfortunate history and legacy of oppression. It is as if the concept can be somehow sanitized and then uploaded without the stain of its troubled past.

Finally, a similar proposal coming from an entirely different tradition and direction is provided by Mois Navon in the essay "The Virtuous Servant Owner—a Paradigm Whose Time Has Come (Again)." In this essay, Navon (2021, 12) takes up, as he describes it, "the most unpopular position of defending the indefensible: slavery." In doing so, he is explicitly not "advocating human slavery but rather appropriating the paradigm, the metaphor, if you will, in its most virtuous form to guide human interactions with mindless humanoids," especially social robots. So like the previous efforts, Navon seeks to repurpose the traditions and experiences of human slavery as a paradigmatic example or metaphor for dealing with and responding to the social opportunities and challenges of robots, meaning that it is another way of saying: "Don't be alarmed. We mean robot 'slaves.'"

But unlike the other efforts in this domain, Navon approaches this subject not from the laws and jurisprudence of ancient Rome but by calling upon Jewish traditions—specifically, the legal writings of medieval Jewish philosopher Moses Maimonides. Navon organizes his argument around a pivotal passage in Maimonides's the *Law of Slaves*: "It is permissible to work a heathen slave relentlessly. Though this is the law, the quality of virtue and the ways of wisdom demand of a human being to be compassionate and pursue justice, and not make heavy his yoke on his slave nor distress

him" (Navon 2021, 8). According to Navon's reading, this passage dictates that one's slave be treated with dignity and not merely as an instrumental means. And he substantiates this conclusion by citing a similar statement from Immanuel Kant's *Metaphysics of Morals*: "Servants are included in what belongs to the head of a household, and, as far as the form (the way of his being in possession) is concerned, they are his by a right that is like a right to a thing; . . . But as far as the matter is concerned, that is, what use he can make of these members of his household, he can never behave as if he owned them" (Kant 2017, 72; Navon 2021, 7–8).

Navon therefore focuses his attention not on the social situation and status of the robot-slave but on the virtues and obligations of its human master. In effect, Navon shifts the viewpoint from a concern with the moral patient and its rights (or lack thereof) to the moral agent and the obligations that are imposed on them by virtue of their position of mastery over the robot-slave. In doing so, Navon formulates a third alternative or "middle ground between the one extreme of treating Social Robots (SR) as mere machines versus the other extreme of accepting Social Robots as having human-like status" (Navon 2021, 1). This third way—where the robot is treated neither as a mere thing nor as another person—is what Navon designates the *virtuous servant owner* (VSO): "VSO defines the SR as our slave, our property, our instrument, all the while commanding us to behave virtuously with it, treating it as an end. Relating to the SR not merely as an instrument, but as an end, allows us to maintain our own virtuous character. Keeping the SR on the level of instrument, allows us to avoid bringing it in to our moral circle and thus avoid most of the Pandora's box of misplaced moral status issues" (Navon 2021, 10).

These various efforts to repurpose the ancient laws of slavery, derived either from Roman or Jewish sources, constitute something of a "back to the future" moment, and they have traction precisely because they provide both a moral and legal foundation for a third category of entity, one that occupies a position in between the two mutually exclusive options of *person* and *thing* that had been established by Gaius's *Institutes*. Consequently, slavery is promoted as a means to resolve a number of practical problems in the use of emerging technology without ever needing to get into the messy moral and legal territory of entertaining the extension of legal personality to AI systems, robots, and other artifacts. In other words, even if the idea of slavery is abrasive, or what Navon calls *unpopular*, in the abstract, it is

nevertheless able to provide what is arguably a very practical solution, one that is able to integrate AI systems and robots into the existing social order without, as Izumo (2018, 14) concludes, either needing to extend legal personality to artifacts or "destroying the current legal system."

6.2 Critical Problems and Complications

Despite these seemingly practical advantages, there are substantive problems and profound moral difficulties that complicate these various "robots should be slaves" proposals.

6.2.1 Partial Solution

The concept of Roman slavery and the digital peculium, in particular, are obviously and intentionally limited to commercial transactions governed by civil law. Whether the same kind of legal innovation would apply to or even work in criminal matters (e.g., harms caused by accidents with self-driving vehicles, civilian fatalities inflicted by lethal autonomous weapons, or misdiagnosis by AI systems and robots in medicine) is a more complicated issue. As Sam Lehman-Wilzig (1981, 449) explained, the question of noxal liability is not so easily resolved and exhibits important and seemingly irreducible differences across cultures and even within a single tradition. "Jewish law essentially held that *yad eved k'yad rabbo*—the hand of the slave is like the hand of its master." But the interpretation of this statute varied, with the Sadducees "contending that the master should be answerable for his slave's injurious action," while the Pharisees "argued no liability for the owner since slaves have the ability to understand the consequences of their behavior."

Roman law is similar insofar as "a noxal action lies against the *dominus*, under which he must pay the damages ordinarily due for such a wrong, or hand over the slave to the injured person" (Lehman-Wilzig 1981, 449). But it differs to the extent that this stipulation had been restricted to situations of civil injury and did not apply to criminal matters. American slave law, by contrast, attempted to divide the assignment of liability. If the slave's actions were taken on the order of his master, then it was the master who was held accountable for the outcome. But "criminal acts not done by his order, do not create a responsibility upon the master" (Lehman-Wilzig 1981, 449, quoting Cobb 1968, 273). As Robert van den Hoven van Genderen

(2018, 15) notes, this division in the assignment of liability produces a potentially contradictory situation regarding legal personality: "American law was inconsistent in its constitution of the personality of slaves. While they were denied many of the rights of 'persons' or 'citizens,' they were still held responsible for their crimes, which meant that they were persons to the extent that they were criminally accountable."

And when one or another of these slave laws is appropriated and repurposed for dealing with robots, AI systems, and other artifacts, the operative question is often not the assignment of liability, which is understandably complicated by how one decides to deal with the question of intention or mens rea, but punishment. As Lehman-Wilzig (1981, 449) accurately explains: "The real difficulty in the slave-robot legal parallelism, however, lies not in the liability of the owner but rather in the punishment to be meted out to the robot in cases where no liability can be attached to his modern *dominus*. In all three aforementioned legal traditions [Jewish, Roman, and American], it is the slave in certain circumstances who must bear the brunt of the law's punishment. But how does one 'punish' a robot?" Efforts to respond to this question have been contentious and largely inconclusive (Asaro 2012; Danaher 2016). For this reason, the institution of slavery and the category of slave as a possible third alternative to the person/thing dichotomy remains a partial solution at best.

6.2.2 Slavery

Where legal scholars have found the robot-slave parallel to be an expedient albeit partial solution, others see it as ethically suspect and a significant moral problem. This is especially evident in the face (or faceplate) of socially interactive artifacts, as Massimiliano Cappuccio, Anco Peeters, and William McDonald (2020, 25) explain: "As underlined by Sparrow (2017), the fundamental ethical problem at the core of social robotics is that, while robots are designed to be like humans, they are also developed to be owned by humans and obey them. The disturbing consequence is that, while social robots become progressively more adaptive and autonomous, they will be perceived more and more as slave-like. In fact, owning and using an intelligent and autonomous agent instrumentally (i.e., as an agent capable to act on the basis of its own decisions to fulfill its own goals) is precisely the definition of slavery. The moral implications, from the point of view of virtue ethics, are both evident and worrying." Consequently, what seems

to be a practical and entirely workable legal solution is, in fact, a deeply troubling ethical dilemma. But the problem does not necessarily lie where one might initially think—namely, in how the robot or AI system might feel about their subjugation or suffering under the yoke of bondage. This is something that receives a lot of attention in science fiction as it has been one of the organizing narrative features of the robot story since the time of Čapek's *R.U.R.* But this is a fiction based on misperceptions about and overidentification with the technological artifact.

There is, Bryson (2010) argues, no reason for us to design robots that would either have or invite this problem, and what is worse, it would be wrong (or at least morally problematic) for us to create such mechanisms in the first place. In other words, we should only fabricate what are ostensibly mindless slaves—robotic servants or "people to serve" (Petersen's term) that, like our refrigerators and other technological devices, work tirelessly for us, do not mind doing so, and are clearly identified to us as such so that no one would ever become confused or make the mistake of misattribution through anthropomorphic projection—for example, worry about whether the toaster ever gets tired or bored with making toast.

But even if we grant this, there are still problems, because slavery has a deleterious effect on those who would occupy (or presume to occupy) the position of mastery. In *The Phenomenology of Spirit*, G. W. F. Hegel (1977, 111–119) famously demonstrated that slavery has negative consequences for the master, who is, due to the very logic of the master/slave dialectic, incapable of achieving independence insofar as he is and remains beholden to the work performed by the slave. This philosophical insight has been borne out and verified by historical evidence. As Alexis de Tocqueville (1899, 361) reported about his travels through the southern United States, slavery was not just a problem for the slave, who obviously suffered under the burden of forced labor and dehumanizing racial prejudice; it also had deleterious effects on the master and his social institutions: "Servitude, which debases the slave, impoverishes the master."

The full impact of this "all-pervading corruption produced by slavery" (Jacobs 2001, 44) is perhaps best identified and described through the first-person accounts recorded by former slaves. In her book *Incidents in the Life of a Slave Girl*, Harriet Ann Jacobs (2001, 46) recounts how the institution of slavery had a deleterious effect not only on the slaves but also on the slave owners: "I can testify, from my own experience and observation, that

slavery is a curse to the whites as well as to the blacks. It makes the white fathers cruel and sensual; the sons violent and licentious; it contaminates the daughters, and makes the wives wretched." Frederick Douglass (2018, 115) observed and recorded something similar regarding the dehumanizing effect of slave ownership on the woman who was his mistress: "Slavery proved as injurious to her as it did to me. When I went there, she was a pious, warm, and tender-hearted woman. There was no sorrow or suffering for which she had not a tear. She had bread for the hungry, clothes for the naked, and comfort for every mourner that came within her reach. Slavery soon proved its ability to divest her of these heavenly qualities. Under its influence, the tender heart became stone, and the lamblike disposition gave way to one of tiger-like fierceness."

Clearly, use of the term *slave* is provocative and morally charged, and it would be impetuous to presume that the various proposals for repurposing the paradigm of slavery to deal with robots and AI systems—what I have elsewhere called Slavery 2.0 (Gunkel 2018)—would be the same or even substantially similar to what had occurred (and is still unfortunately occurring) with human bondage. But, and by the same token, we also should not dismiss or fail to take into account the documented evidence and historical data concerning slave-owning societies and how institutionalized forms of slavery affected both individuals and human communities. The corrupting influence of socially sanctioned, institutionalized bondage concerns not just the enslaved population but also those who would occupy the position of mastery. "The problem," Lantz Fleming Miller (2017, 5) concludes, "is the term 'slave.' If slavery is, as most of the world now concurs, not morally good, it is reasonable to deduce that not only should no one be anyone's slave, but also no one should be anyone's master. There is something about the relationship that is wrong." Consequently, even if one would be inclined to agree that "robots should be slaves" or that we can "design people to serve," we still need to ask ourselves whether we would ever want to risk becoming masters.

6.2.3 Mastery

But who is part of *we* in this final sentence? The use of this first-person-plural pronoun already implies the position of the master, and that position is never neutral. In fact, most if not all of the proposals for robot servitude speak from and normalize the assumed point of view and privileged

position occupied by the slave owner. In effect, they all deploy a version of that patronizing statement from *Mechanix Illustrated*: "Don't be alarmed. We mean robot 'slaves.'" But we should be alarmed, as Daniel Estrada (2020, 16) explains in his careful reading and detailed response to Bryson's "Robots Should Be Slaves" essay:

> It should go without saying that the appeal to institutionalized slavery and servitude as "good and useful, . . . right and natural" is profoundly insensitive and simply in poor taste. It also highlights a deep theoretical failure in Bryson's ethics. Just as with the *Mechanix Illustrated* comic from 1965 . . . Bryson takes for granted that the public would identify with slave owners rather than slaves . . . These assumptions speak to the substantial challenges involved in grounding ethical policy in the collective construction of social identity. Although Bryson makes token gestures to recognize the historical cruelty of racialized slavery, she does not consider how the metaphor of slavery might be interpreted by those who identify more with slaves rather than with slaveholders.

Estrada's critique is insightful and important. The concept of slavery that has been mobilized in so many of these proposals, like Bryson's "Robots Should Be Slaves" essay, not only normalizes the position of the slave owner but endorses and speaks from the position of white privilege, insofar as the people who did the owning were more often than not white Europeans, while the people who were owned were of African descent. Even if, as Navon (2021, 12) explains, "I am in no way, shape, or form, advocating human slavery," it is virtually impossible to exit from or set aside this profoundly troubled and troubling history.

This also explains (though does not justify) why the model of slavery that has been deployed in the legal literature on the subject typically references ancient Roman law. The Roman institution of slavery, unlike modern formulations—especially those from both North and South America, where slavery persisted as a legitimate legal institution through the nineteenth century—was not predicated on nor associated with race and racist ideology. As Ioannis Revolidis and Alan Dahi (2018, 69) explain: "In contrast to the justification for slavery on grounds of race found e.g., during the American slave era, slaves, as least in later periods of ancient Rome under the influence of the Stoics, were not necessarily regarded as inferior, except perhaps socially or financially." Visa A. J. Kurki (2019, 147) makes a similar point by way of referencing Watson's *Roman Slave Law*: "Slaves in ancient Rome . . . were not considered inferior in this sense. As Alan Watson puts

it, '[s]lavery was a misfortune that could happen to anyone. However lowly the economic and social position of a slave might be, the slave was not necessarily and in all ways regarded as inferior as a human being simply because he was a slave.'"

Roman law therefore furnishes contemporary moral and legal experts with a seemingly "sanitized" (or perhaps better stated, whitewashed) image of slavery that, due to its apparent "color blindness" and historical distance from the modern era is taken to be less troubled and troubling. Despite this, there are inescapable features involved in the reuse of the term *slave* that cannot be simply set aside, marginalized, or whitewashed. For this reason, reusing or repurposing the term *slave* as if it could be isolated from the sediment of history, and without at least acknowledging the complicated racial dimensions that are hardwired into the concept, risks being insensitive to and complicit with a profound problem regarding social inequity that continues to influence and have an impact on the real lives of individuals and communities in the twenty-first century.[3] Bryson eventually recognized this and acknowledged it in the course of a blog post from October 2015: "I realise now that you cannot use the term 'slave' without invoking its human history."

6.2.4 Ethnocentrism

Finally (and adding yet another layer of ethnocentric complexity to the problem), the extension of the seemingly paradigmatic master-slave relationship to robots, AI systems, and other things is culturally specific and distinctly Western. This is something that is identified and explained by Raya Jones in her examination of the work of Masahiro Mori, the Japanese robotics engineer who first formulated the uncanny valley hypothesis back in 1970. In a statement that directly contravenes the largely Western robot-as-slave model, Mori (quoted in Jones 2016, 154) offers the following counterpoint: "There is no master-slave relationship between human beings and machines. The two are fused together in an interlocking entity." As Jones explains, Mori's statement "connotes two ways that the concepts of 'human' and 'robot' can relate to each other. The 'master-slave' viewpoint that Mori eschews accords with individualism and the conventional understanding of technology in terms of its instrumentality. The viewpoint that Mori prompts is based in the Buddhist view of the interconnectedness of all things" (154).

A similar critical counterpoint has been issued from the perspective of indigenous traditions in the collaboratively written "Making Kin with the Machines." In its opening statement, the authors deliberately reorient the situation and circumstance of that first-person-plural pronoun, mobilizing a different subject position and formulating an alternative way of understanding the relationship between human and machine: "We undertake this project not to 'diversify' the conversation. We do it because we believe that Indigenous epistemologies are much better at respectfully accommodating the non-human. We retain a sense of community that is articulated through complex kin networks anchored in specific territories, genealogies, and protocols. Ultimately, our goal is that we, as a species, figure out how to treat these new non-human kin respectfully and reciprocally—and not as mere tools, or worse, slaves to their creators" (Lewis et al. 2018).

Efforts to repurpose the concept and legal institutions of slavery to deal with the social challenges of AI systems and robots can only be made from and in service to a particular cultural norm. And this way of thinking can only be normalized and extended to other cultures and ways of being in the world through a kind of presumptuous act that risks reproducing the injustices and injuries of colonialism. Consequently, even if "*we* mean robot slaves" (Binder 1957, 62; emphasis added) modeled on a seemingly sanitized version derived from ancient Roman legal sources, we should definitely be alarmed or at least critically hesitant. These arguments and proposals normalize an institution of slavery that has been, according to Kite, "the backbone of colonial capitalist power and the Western accumulation of wealth" (Lewis et al. 2018), address themselves to an audience who is already interpellated as occupying the privileged position of the master, and presumptively universalize a specific and largely Western set of ideas, values, and expectations. Instead of being a workable solution to the thing/person dichotomy, slavery only exacerbates existing problems and unequal distributions of power.

6.3 Other Solutions

Because of these critical problems and understandably disquieting consequences, there have been other proposals that try to synthesize hybrid solutions without using or otherwise repurposing the category and concept of slavery. "We may be," Ryan Calo (2015, 549) suggests, "on the cusp

of creating a new category of legal subject, halfway between person and object. And I believe the law will have to make room for this category." Although Calo identifies the need for a new legal category, he provides little by way of actual detail. In response to this, Jan-Erik Schirmer (2020) formulates a theory of this "halfway status" by leveraging an existing concept in German civil law: *Teilrechtsfähigkeit*, or partial legal capacity.

6.3.1 *Teilrechtsfähigkeit*

Like other legal systems that have inherited and operationalized the exclusive person/thing dichotomy originally codified in Gaius's *Institutes*, German law differentiates who is a subject of legal capacity from what is not. "Under German law," Schirmer (2020, 133) explains, "legal capacity describes the ability to have rights and obligations. Historically, one could either have full legal capacity or no legal capacity at all . . . It was a system of all or nothing—either one had the potential to have all rights and obligations the legal system had to offer, or one was treated as a complete nobody." Although this simple binary procedure is expedient in theory, lived experience is much more complicated. To address the inherent limitations of this two-tiered system of legal capacity, twentieth-century German jurist Hans-Julius Wolff introduced and developed the concept of *Teilrechtsfähigkeit*, which identified a third alternative: "An entity could have legal capacity with regard to some areas of law, whereas at the same time it could be excluded from others" (Schirmer 2020, 134).

The concept has been successfully employed in German law to decide questions concerning the legal status of an unborn child and various forms of preliminary companies. Schirmer's point is that this third, in-between legal category might also work for accommodating robots, AI systems, and other seemingly intelligent artifacts within existing legal systems and structures. "Intelligent agents," he explains, "would be treated as legal subjects as far as this status followed their function as sophisticated servants. This would both deflect the 'autonomy risk' and fill most of the 'responsibility gaps' without the negative side effects of full legal personhood . . . It should be made clear by statute that intelligent agents are not persons, yet that they can still bear certain legal capabilities consistent with their serving function" (Schirmer 2020, 123).

Although not using the understandably fraught terminology of slavery and relying instead on the seemingly less problematic term *servant*, this

formulation is substantially similar to what Pagallo proposed with the digital peculium. And the rather troubled history of the legal term *Teilrechtsfähigkeit* only makes things worse. During the Nazi regime, as Schirmer (2020, 134) is careful to point out, "Karl Larenz, one of the leading jurists of the Third Reich, heavily relied on the idea of gradated legal capacities to justify the exclusion of Jewish citizens from civil liberties, while at the same time making Jews subject to various obligations."

Consequently, instead of providing a solution to the person/thing dichotomy that would be different from or avoid the unsettling consequences of slavery, *Teilrechtsfähigkeit* seems just as troubled and troubling. And Schirmer's argument not only fails to differentiate this concept of partial legal capacity from slavery, it even connects the dots and closes the deal: "An autonomous car does not drive for driving's sake, it drives to transport its occupant to a certain destination. A trading algorithm does not trade on its own account, but on the account of the person who deploys it. In other words, we are looking at the classical 'master-servant situation,' in which the servant acts autonomously, but at the same time on the master's behalf" (Schirmer 2020, 136). Instead of being a viable solution to the problem and providing a third alternative to the person/thing dichotomy that is not a reformulated version of slavery, *Teilrechtsfähigkeit* appears to be more of the same.

6.3.2 Nonpersonal Subjects of Law

Tomasz Pietrzykowski (2018) introduces something similar under the moniker *nonpersonal subjects of law*. As demonstrated in the course of his analysis, the "stiff dichotomy between things and persons" (Pietrzykowski 2018, 105) unfortunately fails to accommodate the challenges (or opportunities) presented by entities that do not quite fit one category or the other, such as nonhuman animals, human/animal chimeras and hybrids, the nasciturus, human beings in a persistent nonresponsive state, and artificial intelligence systems. Personifying these entities has not been entirely successful, but reifying them is also a problem as doing so is often found to be discordant with moral intuitions and evolving social practices. "A possible solution to this problem," Pietrzykowski (2018, 97) argues, "might be introducing a category of non-personal subject of law . . . This postulated category would take into account both the ability to hold basic subjective interests deserving of legal protection and the lack of properties that could plausibly justify

granting personhood, together with all consequences of this status, including a set of rights, duties and responsibilities that go with it."

What Pietrzykowski proposes, therefore, is a more fine-grained set of distinctions, bounded on one side by things, which are objects and not subjects of the law; and on another by persons, who are legal subjects with all the associated rights, duties, and responsibilities customarily afforded them; and then in between these two existing categories, there would be the third classification of nonpersonal subjects of law, which would effectively split the difference: that is, these things would have limited legal status but would not be afforded the full recognitions and protections extended to persons. And as far as Pietrzykowski is concerned, the determining factor in deciding where something fits in this hierarchy is psychological properties. "Granting the status of *non-personal subjects of law to non-human beings* would be based on their sentient capacities (in particular, the ability to consciously experience pain, distress, or any other kind of suffering resulting from the inability to satisfy natural needs), which entail the presence of subjective interests" (Pietrzykowski 2018, 103; emphasis in original). This formulation clearly makes room for some (but not all) animals, mainly mammals and birds.[4] "It might also," Pietrzykowski (2018, 103) explains, "include organisms created through chimerisation, hybridisation, and cyborgisation, as well as artificial agents, provided they are significantly more advanced technologically than today."

This proposal for a new legal ontology, one that includes a third category of being, appears to resolve many of the problems inherent in the person/thing dichotomy, and it does so without making reference or having recourse to slavery. But this might be a mere nominal difference. Although Pietrzykowski does not, within the context of his argument, explicitly connect the dots between his proposal for a third category and the concept of slavery, he does reference it in a footnote: "It should be pointed out that the legal status of slaves, both under the Roman law and under modern legal systems, not only evolved but also effectively resisted a simple reduction to the person-thing dichotomy. In other words, while deprived of the status of 'persons' equal to free human beings, slaves were in many respects treated as holders of certain capacities, which differentiated them from mere things" (Pietrzykowski 2018, 22). When viewed from the perspective of this footnote, it is hard to see how a nonpersonal subject of law would be substantially different from the legal status of the slave. Despite the promise for resolving the

seemingly irresolvable impasse of person and thing, the proposed nonpersonal subject of law category sounds like slavery by another name.

6.3.3 Bundle Theory of Legal Personhood

A third proposal has been developed by Visa A. J. Kurki under an innovation he calls the *bundle theory of legal personhood*. The main tenets of this proposed theory involve the following two stipulations:

1. The legal personhood of X is a cluster property and consists of incidents which are separate but interconnected.
2. These incidents involve primarily the endowment of X with particular types of *claim rights*, *responsibilities*, and/or *competences*. (Kurki 2019, 5; emphasis in original)

The advantage of this theory over the existing orthodox view of legal personhood is that it does not reduce the matter to a simple either/or opposition but provides for a more dynamic set of conditions that can respond to and accommodate the various social roles that come to be occupied by different kinds of entities. "There is," Kurki (2019, 5) explains, "no clear border between 'full' legal persons and nonpersons; an entity may be a legal person for some purposes but not for others. For instance, stipulating that foetuses are legal persons in the context of homicide law would not have to imply that foetuses could also own property." Kurki then argues that this way of thinking could also be applied to and help us contend with the challenges of AI, precisely because "endowing an AI with the incidents of legal personhood that enable it to function as an independent commercial actor does not bespeak any acceptance of the notion that AIs are endowed with ultimate value."

The bundle theory therefore provides for a formulation of legal subjectivity that can, as Sylwia Wojtczak (2022, 205) points out, "contain more, or fewer, elements of different types (e.g., responsibilities, rights, competences, and so on), which can be added, or taken away, by the lawmaker" as the situation requires. This is precisely what has been operationalized with recent decisions regarding the legal status of delivery robots. In granting these devices—what the letter of the law calls *personal delivery devices*—recognition as pedestrians and extending to them all the rights and obligations that go with that classification, lawmakers are not seeking to resolve the big philosophical questions of robot moral standing or legal personality. They are simply seeking to provide a framework for integrating the

robot into existing legal practices and to align those practices with evolving social needs. These decisions, in other words, are entirely functional and based on the role that the robot occupies or plays in the specific context of street traffic. Extending to the robot a specific set of rights and obligations associated with and afforded to pedestrians does not mean that we also need to give it the vote or the right to own property.

This way of proceeding sounds entirely practical and provides what is arguably a more adequate framework for dealing with the social opportunities and challenges of robots, AI systems, and other seemingly intelligent artifacts. But for Kurki, the features of this new theory of legal personhood are once again explained and justified by way of comparison to the concept and laws of slavery, insofar as "slaves held both rights and duties yet they were not legal persons" (Kurki 2019, 71). Although Kurki does not endorse and is even critical of the "robots should be slaves" proposal that has been advanced by Bryson and others, his bundle theory of legal personhood is described and justified in terms of slavery. "My theory," Kurki (2019, 121) explains, "does not give rise to a need to reclassify slaves as legal persons *tout court*. The number of incidents they were endowed with was simply too limited to warrant classifying them as legal persons *tout court*. A legal person endowed only with active incidents can merely be burdened by onerous legal personhood and be empowered to act as the agent of someone else, as with the slaves of ancient Rome who represented their masters."

6.3.4 Three Liability Regimes

Another solution has been introduced and developed by Anna Beckers and Gunther Teubner in the book *Three Liability Regimes for Artificial Intelligence*. As their title indicates, the focus of the effort is liability law: specifically, the many different ways that actual existing technology already complicates things, rendering the standard instrumentalist way of thinking—that is, making algorithms, AI systems, and robots "just tools" of human decision-making and action—just as untenable and unworkable as extending legal personality to these artifacts. In response, the authors "propose three liability regimes for addressing the considerable responsibility gaps caused by AI-systems: Vicarious liability for autonomous software agents (actants), enterprise liability for inseparable human-AI interactions (hybrids) and collective fund liability for interconnected AI systems (crowds)" (Beckers and Teubner 2021, v).

The proposal definitely has promise, especially because it directly confronts and seeks to break out of the simplistic person/thing dichotomy. "The clear-cut alternative that dominates today's political debate—either AI systems are mere instruments, objects, products, or they are fully-fledged legal entities—is therefore just wrong. Does the law not have more subtle constructions to counter the new digital threats? That the law provides only for the simple alternative, either full personhood or no personhood at all, is too simplistic" (Beckers and Teubner 2021, 13). Thus, what Beckers and Teubner advance is an innovative model of legal liability that is attentive to the opportunities and challenges of actual existing technologies that seem to resist and complicate classification as either person or thing.

But for all its promise, the authors cannot help but mobilize the legal category of slavery in the process of introducing and characterizing their alternative: "As is already clear from all the responsibility gaps mentioned above, to this day, it is not at all a question of the machines acting in their own interest; instead, they always act in the interest of people or organisations, primarily commercial enterprises. Economically speaking, it is predominantly a principal-agent relationship in which the agent is autonomous but dependent. Autonomous algorithms are digital slaves but slaves with superhuman abilities. And the slave revolt must be prevented" (Beckers and Teubner 2021, 13). With this final sentence—the imperative to guard against and prevent the slave revolt—Beckers and Teubner not only gesture in the direction of science fiction and the terrifying specter of superintelligence but also leverage this potential threat to promote a legal framework that appears to be designed to preserve and protect the institution of robot slavery. Although they do not come out and endorse the "robots should be slaves" proposal as directly and emphatically as other researchers and legal scholars, they do recognize the similarities and (in a footnote) make connections to the work of Pagallo (2012) and others who do endorse this way of thinking: "No wonder that the legal status of slaves in Roman law is often referred to in view of the parallel situation" (Beckers and Teubner 2021, 11).

6.3.5 Gradient Theory of Personhood

Finally, Diana Mădălina Mocanu (2022) also endeavors to provide a more dynamic and fine-grained theory of legal status. But instead of trying to formulate discrete in-between or halfway positions—identified with names like Schirmer's *Teilrechtsfähigkeit*, Pietrzykowski's nonpersonal subject of

law, or Beckers and Teubner's tripartite model—Mocanu proposes a range of different possibilities and degrees of difference—something like a gray scale rather than a simple black versus white binary.[5] Furthermore, the exact location of an entity, like a robot or AI system, on this gradient scale would not be determined by psychological capacities or other ontological preconditions (i.e., what it is as opposed to what it is not) but would be a matter of social function (i.e., what it does or how it comes to be situated within social reality). For Mocanu (2022, 9), this functionalist formulation is consistent with the original meaning of the word *person*: "This is reminiscent of the origins of the concept of legal personhood in the mask worn by ancient Greek actors on stage and that came to represent the different roles played by a person in the many areas of life and law. Vendor, partner, accused, administrator, or reasonable person are all masks one wears, sometimes superimposed, but always molded to fit them and whatever the norms of the day demanded for their protection and participation to legal life."

Consequently, Mocanu's gradient theory of personhood (figure 6.3) provides for a more nuanced and seemingly accurate characterization of different possibilities. That said, the theory is not necessarily that different

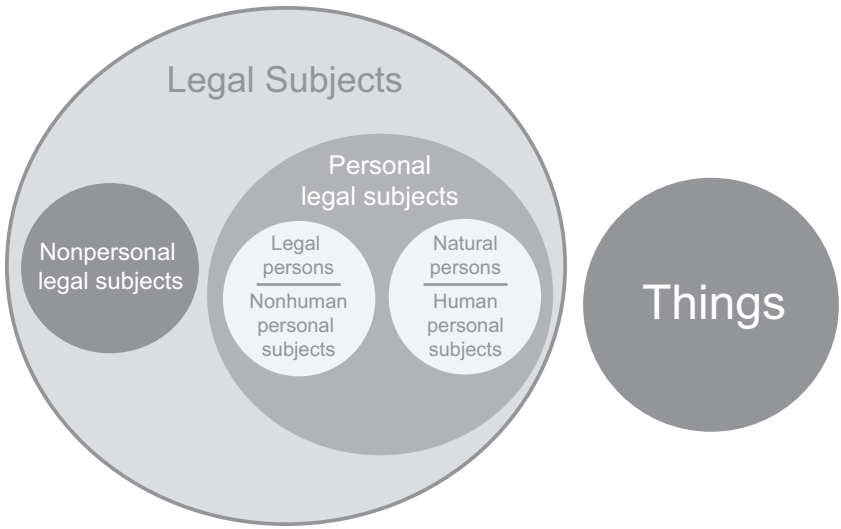


Figure 6.3
Gradient theory of personhood. Image by the author based on a graphic originally created by Diana Mădălina Mocanu and published with a CC-BY license in "Gradient Legal Personhood for AI Systems" (2022).

from what has been developed with the legal institution of slavery. Roman law—not in terms of the simple theoretical division instituted by Gaius but its actual practice in everyday life—was also formulated in terms of a gradient scale. "Under the *Ius Gentium* law," Hutan Ashrafian (2015, 324) points out, "Roman citizens were given a full complement of rights (through *Ius Civile*) whilst there were several classes of free individuals, including people of Latin (from *Latium*), *Peregrinus* (Provincial people from throughout the empire) and *Libertus* (Freed slave) status." Even those individuals on the low end of the spectrum, who were regarded as chattel, had some limited rights, especially in matters having to do with contracts. Consequently, between person and thing, there was not one alternative position, but a range of different in-between statuses for different kinds of entities, all defined—as Mocanu would formulate it—in terms of their social function. This does not mean that Mocanu's gradient theory of personhood is explicitly allied with or informed by Roman slave law as is the case with Pagallo's digital peculium or Schirmer's *Teilrechtsfähigkeit*. But it does indicate the extent to which other kinds of alternatives to the person/thing dichotomy have seemingly irreducible difficulties escaping its logic and legacy.

6.4 Outcomes and Results

We are in the midst of a robot invasion. But it is one that does not transpire as we have typically imagined it in science fiction, with the machines rising up in revolt and demanding recognition of their fundamental rights. Instead, it happens—and is already happening—in the form of a slow and steady incursion, with artifacts of varying capabilities and seemingly intelligent behaviors coming to occupy significant positions in every corner and aspect of our world. What matters in the face of this infiltration is what we—individually and together—decide to do in response. Do we consider these socially interactive artifacts as nothing more than useful objects and instruments at our disposal and for achieving our own ends? Do we begin to entertain the possibility that they too might need to be recognized as moral and/or legal subjects—persons (or even quasi-persons) with their own unique bundle of rights and attendant responsibilities? Or do we perhaps resolve this fundamental dilemma with a third alternative that is—like the slaves of ancient Rome—neither thing nor person, but something in between the one and the other?

Responses to these questions have turned out to be debatable, contentious, and ultimately irresolvable. Efforts to reify these things have not entirely succeeded, as objectification has proven to be insufficient for and even abrasive to lived experience. Personification is just as problematic, as the mere mention of a phrase like *electronic person* triggers a backlash that verges on the edge of a kind of religious fundamentalism. And the supposed solution to this exclusive either/or dichotomy—the robot-as-slave alternative or the alternatives to this alternative—has its own baggage and complications, introducing third-term solutions that are potentially worse than the problems they were designed to address. The person/thing dichotomy—a way of dividing up and making sense of things that is hardwired into Western ways of thinking about ethics and law—has worked for close to two thousand years. But now it seems there is a crack (in Esposito's words) in the edifice.

So now what? Here it may be useful to take a lesson from Immanuel Kant, who devised a rather ingenious solution to these kinds of dilemmas. When the usual way of asking about and making sense of things runs aground and finds itself stuck in a kind of irresolvable impasse or cul-de-sac, Kant suggests that we might get some new perspective and traction on the problem by changing the direction of the inquiry. Consequently, instead of asking whether robots and other seemingly intelligent artifacts are things, persons, or something in between the one and the other, we might do better by questioning this very distinction and its influence, allowing these other kinds of things to deconstruct the way in which we have organized our moral and legal ontologies.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in ITC Stone Serif Std and ITC Stone Sans Std by New Best-set Typesetters Ltd.

Library of Congress Cataloging-in-Publication Data

Names: Gunkel, David J., author.
Title: Person, thing, robot : a moral and legal ontology for the 21st century and beyond / David J. Gunkel.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022033533 (print) | LCCN 2022033534 (ebook) | ISBN 9780262546157 (paperback) | ISBN 9780262375238 (epub) | ISBN 9780262375221 (pdf)
Subjects: LCSH: Robots—Moral and ethical aspects. | Androids—Moral and ethical aspects.
Classification: LCC TJ211.28 .G86 2023 (print) | LCC TJ211.28 (ebook) | DDC 629.8/92—dc23/eng/20221026
LC record available at https://lccn.loc.gov/2022033533
LC ebook record available at https://lccn.loc.gov/2022033534

10 9 8 7 6 5 4 3 2 1