California Law Review Circuit

The Path of Robotics Law

Jack M. Balkin[*]

INTRODUCTION

What are the key problems that robotics presents for law? To answer this question, Ryan Calo argues that we should make an analogy to cyberlaw. First, we should identify the "essential qualities" of the new technology, and then we should ask how the law should respond to the problems posed by those essential qualities.[1]

Calo's account of the problems that robotics present for law is just terrific, and I believe it is destined to be the starting point for much future research in the area. This Essay builds on his ideas and takes them in a somewhat different direction.

First, I draw different lessons than Calo from the history of cyberlaw and from encounters between law and new technology. I do not think it is helpful to speak in terms of "essential qualities" of a new technology that we can then apply to law. On the contrary, we should try not to think about characteristics of technology as if these features were independent of how people use technology in their lives and in their social relations with others. Because the use of technology in social life evolves, and because people continually find new ways to employ technology for good or for ill, it may be unhelpful to freeze certain features of use at a particular moment and label them "essential."

Second, I describe the problems posed by robotics for law a little differently than Calo does, in part because I do not distinguish sharply between robots and artificial intelligence (AI) agents. As innovation proceeds, the

---

[*]   Knight Professor of Constitutional Law and the First Amendment, Yale Law School.
1.   Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 514–15 (2015).

distinction between these two kinds of technologies may be far less important to the law than it seems at present; we do not yet know whether the boundaries between these two technologies will increasingly blur (as I expect) or further differentiate.

Instead, I will emphasize two key problems that robotics and artificial intelligence present for law. The first is how to distribute human rights and responsibilities that arise from the actions of nonhumans. Human beings will use robots and AI entities to create new things and to violate the legally protected interests of other human beings. AI agents will churn out novels; but they will also spy on people and defame them. Robots will invent new things; but they will also break people's limbs. Calo suggests, correctly, I think, that we are still a long way from treating robots and AI agents as self-conscious rights-bearing or responsibility-bearing entities.[2] Therefore the key question for law—at least at this early point in history—is how to allocate rights and duties among *human beings* when robots and AI entities create benefits or cause injuries.

The second problem is a generalization of the first. I call it the "substitution effect"—the fact that people will substitute robots and AI agents for living things, and especially for humans. But they will do so only in certain ways and only for certain purposes. This substitution is likely to be contextual, unstable, and often opportunistic. The problem of substitution touches many different areas of law, and it promises to confound us for a very long time.

I.
THE LESSONS OF CYBERLAW

Oliver Wendell Holmes Jr.'s famous 1897 lecture, *The Path of the Law*,[3] argued that because law is produced by the cumulative forces of social life, we should view law from the standpoint of its social function and practical use. Holmes offered a proto-Realist manifesto against formalism and the belief in essential features of legal concepts. All American lawyers, Calo and myself included, are intellectual descendants of Holmes's pragmatism. Rejecting essentialism in law helps us focus on law's basis in social relations and law's role in maintaining—and occasionally mystifying or apologizing for— relationships of authority and power. We should adopt a similar anti-formalism when we think about technology and law's encounters with innovation.

When we consider how a new technology affects law, our focus should not be on what is *essential* about the technology but on what features of social life the technology makes newly *salient*. What problems does a new technology place in the foreground that were previously underemphasized or deemed less

---

2. *See id.* at 517 (taking a "conservative" view of robotic capacities in the near future), 539 ("Little is gained, and much is arguably lost, by pretending contemporary robots exhibit anything like intent.").

3. Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457 (1897).

important? What aspects of human activity or of the human condition does a technological change foreground, emphasize, or problematize? And once technology moves a certain problem or a certain feature of our lives from the background to the foreground of our concern, what are the consequences for human freedom?[4]

The features that are most salient about our technology may depend on how people come to use it. Often people use technology in ways its designers did not foresee or intend. That is especially so for what Jonathan Zittrain calls "generative" technologies,[5] which offer multiple sites and possibilities for innovation. What seems particularly important and salient about technology changes over time as people work with and through new technology. Moreover, technology, like law, mediates social relations between human beings—including relations of power and control. Because those relations are always evolving, our assessment of what is most interesting or worrisome about a technology may change too.

Today, people carry microcomputers in their purses and pockets called "cell phones." Thirty years ago people might have argued that an essential characteristic of a cell phone was its ability to make a phone call outside of one's home. But this feature of cell phones is by no means the primary way that people use them today. Instead, people's habits developed alongside developments in mobile technology. What people use their phones for now— through a kaleidoscope of different applications—makes the very word "phone" seem like a relic of a past age.

In 1991, before widespread adoption of the World Wide Web, the Internet was used mostly for e-mail and for exchanging files through protocols like Gopher. (Search functionality was primitive and Google had yet to be invented.) Then, one might have imagined that the Internet's "essential" features were its abilities to cross jurisdictional lines at will, to send digital information quickly and cheaply, and to facilitate anonymous communication. But this is because, at that point in history, that is how people imagined human beings would use the technology and the kinds of social relations of power, action, and vulnerability that these uses would engender. The characteristics of a new technology, in short, are partly the product of current use and partly the work of human imagination about potential affordances and opportunities, dangers and threats.

By 1999, it was clearer that states could control features of Internet traffic and that the degree of truly anonymous communication the Internet afforded was overstated. Dreams of multiple cyberworlds free from the reach of territorial governments had given way to increasing concerns about filtering

---

4.   Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 2 (2004).

5.   JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET: AND HOW TO STOP IT 67 (2008).

and surveillance. The first edition of Lawrence Lessig's book, *Code*,[6] emphasized how much the Internet had changed between 1995 and 1999. A key theme was that these changes showed that the Internet did not have to be designed in any particular way, and that it could become either characteristically free or controlled.[7] Similarly, as Yochai Benkler pointed out, the Internet is not a single medium of communication—whether one-to-many or many-to-one or many-to-many—but rather a platform for creating multiple media with many different characteristics, affordances, and opportunities for social action.[8] By the time Benkler wrote *The Wealth of Networks* in 2006 and Clay Shirky published *Here Comes Everybody* in 2008,[9] the Internet seemed to have been transformed once again. Now, its most prominent feature was its ability to lower the cost of organization, to facilitate crowdsourcing and open-source projects, and to undercut professional norms of information production in areas ranging from music to journalism to science. Calo's own account of the "essential" characteristics of the Internet[10] is a retrospective view that builds on more than two decades of social and technological innovation, rather than a timeless account of inherent properties.

The Internet of 1991 is so unlike the Internet of 2015 that it is almost hard to imagine. Although there are family resemblances between the two, the differences are vast, due to the amazing ingenuity of countless human beings who innovated in social and economic relations as much as in hardware and software.

We might say, then, that the most important lesson of cyberlaw for robotics is the need to attend to the relationships between affordance and imagination, between tools and relations of power, between technological substrate and social use. The characteristics of robotics that currently draw our attention, that disturb our legal complacency, that create puzzles for law and opportunities for social innovation, will not be the last puzzles and opportunities we face.

Indeed, what lawyers call "technology" is usually a shorthand for something far more complex. When we talk about "technology," we are really talking about (1) how people interact with new inventions and (2) how people

---

6.   LAWRENCE LESSIG, CODE: AND OTHER LAWS OF CYBERSPACE 27–28, 205, 216 (1st ed. 1999).

7.   *Id.* at 25 ("There is no single way that the Net *has* to be; no single architecture defines the nature of the Net.").

8.   *See* YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM 370 (2006) ("[T]he Internet allows for a radically more diverse suite of communications models than any of the twentieth-century systems permitted.").

9.   CLAY SHIRKY, HERE COMES EVERYBODY: THE POWER OF ORGANIZING WITHOUT ORGANIZATIONS (2008).

10.   Calo, *supra* note 1, at 518–25. Calo suggests that over time the key legal problems of the Internet have become clearer and more stable. *Id.* at 525–27. If that is true—and I am somewhat more skeptical than Calo—it is because, in hindsight, we more clearly see the path we have travelled through more than twenty years of continuous interaction between technological and social innovation.

interact with other people using those new inventions or presupposing those new inventions. What we call the effects of technology are not so much features of *things* as they are features of *social relations* that employ those things. These social relations include relationships of power and authority, forms of human organization and production, and features of human sexuality, association, and family life.

Similarly, innovation in technology is not just innovation of tools and techniques; it may also involve innovation of economic, social, and legal relations. As we innovate socially and economically, what appears most salient and important about our tools may also change. If the Internet of 1991 looks radically different from that of 2015, it is not only because our tools have evolved; it is also because how people live, work, and contend with each other has evolved in interaction with those tools. As our world fills with robotic and AI technologies, our lives and relationships of social, political, and economic power will also change, posing new and unexpected challenges for law.

II.
WHAT DO ROBOTICS AND ARTIFICIAL INTELLIGENCE MAKE SALIENT?

A.   Let's Get Physical

Calo explains that robots create new problems for law because robots can cause physical harm rather than just economic or emotional harm.[11] Calo calls this feature "embodiment,"[12] but that term may be a bit misleading. Computers have always been embodied in physical form—think of the huge IBM mainframes of the past. Moreover, computers have always had the ability to cause physical harm because of their physical embodiment—for example, a laptop can cause physical injury if it is thrown, or if it is dropped on a person's foot and breaks a toe.

Rather, Calo's point is that robots can cause physical harm *because of their programming*, or more precisely, the cumulative effect of their hardware, operating system, and software.[13] These allow a robot to interact with its environment and have physical effects on the world. So although Calo's concern may seem to be primarily about the capacity to cause physical harm, the deeper issue is that harm—whether physical, emotional, or monetary—is caused by programmability and interactivity. The fact that robots have bodies plays a different role in what Calo calls "social valence"—the notion, among other things, that people tend to treat interactive moving objects as if they are alive.[14] But here again, what is important is not that robots have a physical

---

11.   *Id.* at 515, 542.
12.   *Id.* at 532.
13.   *Id.* at 533–35.
14.   *Id.* at 532–33.

form—so do toasters—but that their interactivity creates particular social cues in human beings.

A robot's ability to cause physical injury is not really an "essential" characteristic of robotic technology. It is a particularly *salient* feature of robotics for lawyers, but the reason that it is salient is that, as Calo explains, from the early days of the Internet, judges decided to adapt existing tort law doctrines to limit liability for purely economic losses caused by new digital technologies.[15] Similarly, Congress created a series of legislative immunities for online service providers that limited liability for copyright infringement or harmful communication that used digital networks.[16] The problem of physical injury is not simply a feature of essential characteristics of a technology. Rather, it arises from the way that a new technology interacts with a social and legal world already in place.

A new technology enters into a social world already in motion, with an existing set of assumptions and expectations about what is likely and unlikely, possible and impossible, an existing set of norms about social life, and a set of paradigmatic examples about how things work and what we should do in response to problems. From law's standpoint, we might call these assumptions and expectations about how the world works the *scene of regulation*.[17] The new technology disrupts the existing scene of regulation, leading various actors to scramble over how the technology will and should be used. As people scramble and contend with each other over the technology, they innovate—not only technologically, but also socially, economically, and legally—leading to new problems for law. Instead of saying that law is responding to essential features of new technology, it might be better to say that social struggles over the use of new technology are being inserted into existing features of law, disrupting expectations about how to categorize situations.

Classifying the capacity of programmable entities to create physical harm as an "essential characteristic" of robotics has another disadvantage. It may lead us to neglect the diversity of systems that employ artificial intelligence and self-learning. Indeed, as Calo points out, there is a continuum between "robots" and "artificial intelligence."[18] That is because, like the Internet itself, robots and other interactive entities do not have to be designed in any particular way. And because there is a continuum of potential designs and a variety of different potential uses, there is also a continuum of potential effects that these new technologies can have. Self-learning algorithms can raise or lower temperatures

---

15.   *Id.* at 536–37.

16.   *Id.*; Online Copyright Infringement Liability Limitation Act, 17 U.S.C. § 512(a)(1) (2012); Communications Decency Act of 1996, 47 U.S.C. § 230 (2012).

17.   *Cf.* Jack M. Balkin & Reva B. Siegel, *Principles, Practices, and Social Movements*, 154 U. PA. L. REV. 927, 931 (2006) (defining a "regulatory scene" as "a paradigmatic set of problems" that a legal "principle is supposed to resolve and a particular set of paradigmatic practices that the principle legitimates or delegitimates").

18.   Calo, *supra* note 1, at 530–32.

in a house, turn on appliances, lock or unlock gates, and notify security services. Algorithms can buy and sell securities; they can create holographic projections that look and act like people; they can threaten, entertain, copy, defame, defraud, warn, console, or seduce. These various effects straddle the lines between the physical, the economic, the social, and the emotional.

We may be misled if we insist on too sharp a distinction between robotics and AI systems, because we do not yet know all the ways that technology will be developed and deployed. Similarly, we may also be misled if, in order to preserve the theory of "essential" characteristics, we try to shoehorn all of these effects and harms into "physical" effects and harms. A robotic cat designed as a companion for an elderly person to pet and care for engages the physical and the emotional. A smart home that controls temperature, orders our groceries, and briefs us on our day combines the physical, the economic, and the social. We should be attentive to the diversity of uses that our technologies will have, and therefore to the diversity of harms and threats that these technologies might pose.

We might identify many different features of a technology as its key or essential characteristics, but the real issue is always *why* we care about them. How we define the central features of a technology depends on what our definition is *for*, and the purpose it serves in our particular area of inquiry. If we are engineers trying to solve a design problem, we will focus on different aspects of the technology than if we are lawyers trying to solve a problem of legal rights and liabilities. Not surprisingly, lawyers tend to view technology through the lens of existing legal doctrines and present policy concerns. Lawyers may be interested—and should be interested—in how engineers and computer scientists define their terms; but lawyers' interest flows from their ability to relate these ideas to legal and policy problems.

B.   Code is Lawless

Both robots and AI agents create problems for law because one cannot always predict what they will do when they interact with their environment. Calo calls this feature "emergence"[19] or "emergent behavior."[20] I agree with him that, at least at this point in history, "emergence" is a more helpful term than "autonomy," which raises a host of difficult questions about the status of artificial intelligence. The problem of "emergence," however, cross-cuts with the problem of physical harm. Several of Calo's examples—flash trading, algorithmic speech, and automated copyright or patent infringement—involve emergence but do not threaten physical injury.[21] Nevertheless, some of these injuries—for example, the destabilization of securities markets—can be

---

19.   *Id.* at 532.
20.   *Id.* at 515.
21.   *Id.* at 541–43.

catastrophic even if they do not involve physical harm. Surely these kinds of injuries are just as important to law's encounter with new technology, but they further undermine the assumption that a capacity to cause physical injury is an "essential" characteristic of the technologies we are interested in.

By contrast, the idea of emergent behavior may seem far more "essential" to robotics, especially when we consider robots together with AI systems generally. But the importance of this feature, too, is produced by the interaction of new technologies with existing law. If law had easy or obvious solutions to injuries by robots and AI systems, we would not consider the fact that robots and AI systems make decisions for themselves to be particularly distinctive or salient in the eyes of the law, even though this fact might be important for technological, moral, or philosophical reasons. As Calo suggests, however, emergent behavior by robots and AI systems does create new difficulties for existing legal doctrine.[22] That is why lawyers care about it.

The problem of emergence is the problem of who we will hold responsible for what code does. Lawrence Lessig's famous dictum that "Code is Law" argued that combinations of computer hardware and software, like other modalities of regulation, could constrain and direct human behavior.[23] Emergence presents the converse problem: self-learning systems may be neither predictable nor constrained by human expectations about proper behavior. Code is lawless.

We can divide this concern into two problems. First, we must assign responsibility for injuries—whether physical, economic, or emotional—caused by interactive and/or self-learning systems. We might hold many different potential actors liable, including the owner, operator, retailer, hardware designer, operating system designer, or programmer(s), to name only a few possibilities.

What degree of fault should we require of these potential defendants? The easy case, and the one for which the law is already well prepared, involves a defendant who programs a robot to harm another. But in most cases, it will be difficult to show either a deliberate intent to harm or knowledge that harm will occur. One might then turn to theories of negligence. But although the risk of *some* kind of injury at *some point* in the future is foreseeable whenever one introduces a new technology, how and when an injury occurs may not be particularly foreseeable to each of the potential defendants in the above list. If the law hopes to assign responsibility to humans and corporations, injuries by robotic and AI systems may strain traditional concepts of foreseeability. Liability without fault is a traditional solution, but it may stifle innovation in a developing area, and it may not be an appropriate solution in the context of criminal law.

---

22.   *Id.* at 543.
23.   LESSIG, *supra* note 6, at 6, 20, 88–89.

A second problem—of causal responsibility—follows from the first. The programming and algorithms used by robots and AI entities will likely be the work of many hands. The example of digital networks is instructive. Jonathan Zittrain's concept of generativity explains the Internet's seemingly inexhaustible capacity for innovation in terms of multiple layers of potential variation.[24] There is a hardware layer through which digital content travels, a protocol layer that creates a common language for exchange, and successive layers of software built one upon the other.[25] And there is also a "social layer"—the social activities, customs, and organizations that grow up around generative technologies and find interesting and novel ways to use and alter them.[26] Innovation may occur in each layer, combining the work of many different people.

For example, at the software layer, open source projects take advantage of the Internet's ability to lower the costs of organization. These projects—and many commercial projects as well—are often perpetually changing through successive updates and fixes. Software programs may also serve as platforms for apps, plug-ins, and modifications of infinite variety produced by third parties.

Robots and AI algorithms build on these practices of innovation; we can understand them in terms of multiple layers of hardware, protocol, and software offering multiple opportunities for innovation and alteration by multiple actors. The more opportunities for innovation and alteration, however, the more difficult it will be to locate and demonstrate responsibility for emergent behavior that harms another.

Software—especially mature and complex software—is likely to have bugs or produce unpredictable results. Bugs may be difficult to spot and may develop through the combination of multiple modifications and additions. It may be fiendishly difficult to affix responsibility for bugs that emerge from layers of software development by many hands. And to the extent that robots and AI programs learn how to modify their own code, the questions of responsibility become even more diffuse.

The problem of security is the flip side of the problem of causal responsibility: the more opportunities for innovation, the more possible targets for hacking. Here, Lessig's famous idea of regulation by code turns into its opposite. Instead of code as a law that regulates humans, code features a kind of lawlessness that escapes human regulation. As James Grimmelmann has pointed out, it is characteristic of much regulation by software that it is hackable and not robust.[27] Once a software system is hacked it may fail

---

24.   ZITTRAIN, *supra* note 5, at 67.
25.   *Id.* Zittrain calls these the "physical" layer, the "application" layer, and the "content" layer. I am using slightly different terminology.
26.   *Id.*
27.   James Grimmelmann, *Regulation by Software*, 114 YALE L.J. 1719, 1741–43 (2005).

instantaneously in ways that are not immediately obvious; and the system may be unable to recover on its own without human intervention.[28]

To make matters worse, many robots and AI systems will probably be continually connected to the Internet and will continually take in new information and new programming from multiple sources. Self-driving cars, for example, could be designed as part of a giant network of interactive robots, constantly sending each other information about local driving conditions. Regular updates to the operating system might be downloaded to each car without the end-user's knowledge.[29] Indeed, we should expect that some of the most useful and widely employed robotic and AI systems will be connected to the Internet cloud. This means that these systems will not be self-contained entities, but will continually be updated by communication with other robots and AI entities, as well as centralized and decentralized sources of information. Quite apart from the security problems that cloud robotics presents, it also complicates and diffuses responsibility for accidents.

Calo explains that these problems follow from the "promiscuity" of data that is characteristic of digital networks generally.[30] The metaphor of "promiscuity" suggests that nodes on the network are continually sharing information with many other nodes (like a person who has intercourse with many different people). The metaphor of "promiscuity" also suggests vulnerability to infection (by analogy to venereal disease), hacking, and harm. Calo connects the promiscuity of data to the problem of embodiment—that is, the capacity to cause physical harm.[31] That is because promiscuous sharing of data and ease of reprogramming wasn't such a problem when the only harm that resulted was nonphysical.

I would look at the issue differently. I would associate the interconnected nature of the Internet, digital generativity, and the multiple layers of innovation with the problem of emergence. That is because from the standpoint of law—as opposed to the standpoint of engineering—the problem posed by emergence is the problem of assigning responsibility for the unpredictable behavior of robots and AI systems.

Calo emphasizes that this unpredictability comes from the complexity of algorithms, including those that learn from experience.[32] That is certainly an important cause of the problem for law, but it is not the only one. Multiple layers of innovation, the generativity of digital systems, and the easy flow of data are also reasons why we may be unable to predict what self-learning or

---

28.   *Id.* at 1742.
29.   Cf. Jack Boeglin, *The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation*, 17 YALE J.L. & TECH. 171, 181 (2015) (describing "communicative" automated vehicles that dynamically download data from other vehicles).
30.   Calo, *supra* note 1, at 515, 519, 532–33.
31.   *Id.* at 533–35.
32.   *Id.*

interactive systems will do once we set them loose on the world. This is yet another reason why "emergence" is a better term than "autonomy." A self-learning system is not made more *autonomous* because of the fact that it is structured in layers, generative, modular, buggy, and the work of many hands. Rather, these features are related to the reasons why the law cares about emergence: they enhance the possibility that self-learning systems will solve problems and perform tasks in ways that we will not foresee or expect.

The problem of emergence works in the opposite direction as well—instead of *harms*, emergent behaviors may create multiple social *benefits*. For example, robots and AI systems will create new inventions and literary works. The question is who will enjoy the intellectual property rights, or whether, instead, the creations will enter the public domain.[33] Search algorithms produce speech in the form of rankings; they may also automatically fill in or complete search queries. Should we treat this expression as part of the speech of the search engine company, and therefore protect it from regulation under the First Amendment?[34]

We are likely to see considerable opportunism in the legal arguments that people make about robots and AI systems. People will claim credit (and legal protection) for whatever benefits their robots create, while denying responsibility whenever their robots destroy property, infringe copyright, make threats, or engage in defamation. People will downplay emergence when the system produces benefits that they might capture for themselves and emphasize emergence when the system causes harm. It is a variation on the old saying that "success has a thousand fathers; failure is an orphan." The goal for law is to meet this predictable opportunism, and to come up with a rational allocation of benefits and responsibilities for both robotic creations and robotic injuries.

III.

THE SUBSTITUTION EFFECT

The third feature of robotics that Calo calls essential is the tendency of human beings to respond to robots as if they are interacting with people. He calls this tendency "social valence,"[35] and argues that "to a greater degree than perhaps any technology in history, robots have a social meaning to people."[36]

---

33. Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5; Ralph D. Clifford, *Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand Up?*, 71 TUL. L. REV. 1675 (1997); Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 1185 U. PITT. L. REV. 47 (1985).

34. Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495 (2013); Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445 (2013); Eugene Volokh & Donald M. Falk, *Google: First Amendment Protection for Search Engine Search Results*, 8 J.L. ECON. & POL'Y 883 (2012); Frank Pasquale, *Search as Speech: Two Scenarios*, CONCURRING OPINIONS (May 29, 2012, 9:53 AM), http://www.concurringopinions.com/archives/2012/05/search-as-speech-two-scenarios.html.

35. Calo, *supra* note 1, at 532, 545–46.

36. *Id.* at 546.

Of course, people do not necessarily respond to robots as if they were people. They may respond to them as if they were animals. Consider a robotic dog that one trains to do tricks, or a robotic seal designed for elderly patients to pet. What Calo is describing might be either anthropomorphism (ascribing human features) or zoomorphism (ascribing animal features). The more general point, however, is that human beings project agency on non-living things. Humans may also project emotions, feelings of pleasure and pain, the capacity to form relationships with others, and the capacity to care for others and be cared by them in turn. The projection of human or animal emotions onto inanimate objects is as old as history itself. People hear the wind howl and the ocean roar; they project agency and loyalty onto their ships and cars. The projection of humanity onto what is not human is the reflection of the self on the outside world.

It is worth noting, once again, that this tendency is not unique to robotics; it is also true of AI systems. Spike Jonze's 2013 movie *Her*[37] is about a man who falls in love with an operating system, not a robot. Robots may cause people to see them as alive because they move; but AI systems may cause people to see them as alive because they speak. And if the concern is anthropomorphism, people have associated the power of speech with humanity far more than the power of motion.[38]

The first two characteristics of robotics that Calo identified—the capacity to cause physical harm and emergent behavior—create obvious problems for assigning liability in tort and criminal law. That is why they are particularly interesting to lawyers. What Calo calls "social valence," however, is a far more complex phenomenon. It is not limited to the question of legal liability but concerns every way that robots and AI agents might intervene in social relations. Not surprisingly, Calo offers a diverse array of possible consequences. For example, he points out (1) that the more anthropomorphic a robot, the more people will assign blame to the robot rather than to a person using the robot;[39] (2) that the presence of robots in a system of surveillance heightens the subjective sense that one is being observed;[40] (3) that human beings will take greater risks to preserve the integrity of anthropomorphic robots than they would for things designated as tools;[41] and (4) that human beings may suffer distinctive emotional harms for the loss of robotic companions.[42] Finally, building on the work of Kate Darling, Calo suggests that anthropomorphic robots pose problems of moral harm: people who

---

37.   HER (Annapurna Pictures 2013).
38.   *See, e.g.*, JOHN HEATH, THE TALKING GREEKS: SPEECH, ANIMALS AND THE OTHER IN HOMER, AESCHYLUS, AND PLATO 1 (2005) (arguing that the Ancient Greeks saw the power of speech as the most important difference between human and animal natures).
39.   Calo, *supra* note 1, at 547–48.
40.   *Id.* at 547.
41.   *Id.* at 548.
42.   *Id.*

mistreat anthropomorphic robots may coarsen or degrade themselves in the same way as people who abuse people and animals, even if no third party is actually harmed.[43]

Calo's brief discussion scratches only the surface of what is at stake, and "social valence" may not be the best way of describing it. All of his examples revolve around a basic problem produced by the introduction of robotics and artificial intelligence into society. The problem is not that people *confuse* robots for living things, for usually they do not. Rather, the problem is that, through their interactions with robots and AI systems, people are willing to *substitute* them for animals or human beings in certain contexts and for certain purposes. Call this the *substitution effect*. People cause an entity to stand in for a human or animal and they treat the entity as such—but only in certain ways. In other words, people treat robots and AI agents as *special-purpose animals* or as *special-purpose human beings*. Calo puts it well when he speaks of "a new category of legal subject halfway between person and object."[44] The placement is "halfway" because the assignment of status may be incomplete, contextual, unstable, and, above all, opportunistic. People may treat the robot as a person (or animal) for some purposes and as an object for others.

I call this a substitution effect because in many cases we do not regard a substitute as fully identical to the thing for which it substitutes. Rather, it is only equivalent provisionally, in certain contexts or for certain purposes, and people often reserve the right to reject the asserted identity when it suits them. The substitute teacher in a fourth grade classroom may lack the students' full respect, and probably doesn't enjoy the benefits of the regular instructor's pension plan. Despite what the manufacturers of substitutes for butter, cream, and sugar tell us, we don't treat these dietary substitutes as the genuine article. Rather, we use them to avoid the calories or the fat associated with the real thing. That is, people use the substitute to try to get the best of both worlds— the students get taught about fractions at lower cost; we get the satisfaction of butter with half the calories.

Projecting human emotions, feelings, and goals onto inanimate objects is a special case of substitution. Our imagination makes the robot stand in the place of a human being or animal in some respect. This is the original meaning of a "substitute," which comes from the Latin *substituere*, derived from *sub+statuere*, meaning to set up under or stand in the place of another.[45] A substitute is something set up or placed under something to replace or displace it.

---

43.   *Id.* at 548–49 (citing Kate Darling, *Extending Legal Rights to Social Robots*, *in* ROBOT LAW (Ryan Calo et al. eds.) (forthcoming 2015), *available at* http://ssrn.com/abstract=2044797).
44.   *Id.* at 549.
45.   *See* MERRIAM WEBSTER ONLINE DICTIONARY, http://www.merriam-webster.com/ dictionary/substitute (providing etymology of *substitute*).

The substitution of robot for living thing may be innocent, emotional, almost instinctual. The patient who blames a surgical robot for a botched procedure projects a partial humanity—and hence responsibility—onto the technology. The soldier who mourns the loss of his bomb-disarming robot projects onto the robot human qualities of comradeship, courage, and commitment to fellow soldiers. When a companion robot who operates in our home sends personal data about us to a corporation, we feel betrayed, when we would never think that a camera and a microphone could betray us.

Most of the forms of substitution that Calo describes are not deliberate—they may simply reflect how human beings react to moving objects and speaking programs. But equally interesting forms of substitution are deliberate and instrumental. A government might replace human soldiers with robots because the latter have no families and don't come home in body bags; their wanton destruction is less likely to undermine political support for a war. Corporations may substitute robots for workers because robots won't unionize, won't need coffee breaks, and don't suffer from alcohol abuse, emotional problems, or other causes of subpar performance and absenteeism. This practice of substitution, unlike the emotional or instinctual substitution described above, is conscious and purposive. Our choice to use robots is like our choice to use margarine or nondairy creamer—it seems to offer all of the benefits of humans with none of the costs.

The law might also deliberately treat robots as if they were living agents for practical reasons or for reasons of public policy. For example, we might protect robots from certain forms of abuse because we fear that people who abuse robots might also abuse people, children, or animals. We might impute mens rea to a robot or algorithm as a legal fiction in order to make it easier to apply the law of respondeat superior, criminal concepts of aiding and abetting, or accomplice liability to owners, operators, or designers. Similarly, courts might treat AI-produced art as a "work for hire" in order to minimize changes to existing copyright law. When we employ legal fictions of this kind, we substitute robot for human to allow the law to function effectively in the face of the legal enigmas posed by emergent behavior. Or we might adopt legal fictions to keep existing legal doctrines working provisionally until we can produce more thorough and coherent reforms.

I have just described two kinds of substitution—unconscious or emotional, and purposive or instrumental. In fact, they are two sides of a coin. A substitute *serves* in the place of another—like a substitute teacher—or is *used* in the place of another—as when we substitute margarine for butter in a recipe. Notions of service and use apply both in emotional substitution and practical substitution. We make robots serve our emotions and our practical needs; we use them to fulfill our desires and our projects. As in Hegel's famous dialectic of master and slave, we may become dependent on what we use and

on what serves us.[46] Robots and AI agents may perform tasks that we no longer can (or are willing to) perform for ourselves; they may offer us emotional succor that we find we can no longer do without.

A key feature of robotic substitution is that it is *partial*. Robots and AI entities take on particular aspects and capacities of persons—for example, their capacity to labor, their capacity to fight and kill, their capacity to nurture, their capacity to create, their capacity to console, entertain, or sexually satisfy. Through using robots and AI agents, human characteristics and skills can be segmented, isolated, and dispersed. This is what I mean when I say that robots and AI agents operate as "special-purpose human beings"; they are agents for a particular reason or function, straddling the line between selves and tools, or persons and instruments.

At first glance, the notion of a "substitute"—and only a partial substitute at that—might suggest lack or inferiority. Margarine is not as healthy as butter; the apprentice mechanic who shows up is less skilled than the one who couldn't make the scheduled appointment. But substitution may also involve abundance and superiority. Substitution not only promises the possibility of greater benefits with lower costs, but it may offer skills and performances better than humans can offer. In such cases, the substitute promises to be super-human. The robot companion promises to better cater to our emotional needs, the AI trading program engages in arbitrage faster and more efficiently than the floor trader, the robot army mows down the enemy more efficiently, and so on. In fact, it is the very partiality of the substitution that makes it superior. The robot caretaker doesn't tire of attending to our needs, and the robot soldier doesn't suffer from posttraumatic stress disorder, because neither of them is fully human.

I began this Essay by noting that technology concerns not merely the relationship of persons to things but rather the social relationships between people that are mediated by things. When the "things" at issue are substitutes for people, or special-purpose people, the mediation of social relationships between human beings is even further obscured. Our interactions with robots and AI systems are interactions with the people who are deploying these new technologies, even when we do not realize it.

CONCLUSION

In this Essay I've argued against essentialism in law's encounter with technology, advocating instead that we should always keep the social aspects of technology in mind. Because we innovate in social relations along with technology, we cannot always tell what will be most important about

---

46. GEORG WILLIAM FRIEDRICH HEGEL, THE PHENOMENOLOGY OF SPIRIT 117 (Arnold Vincent Miller trans., 1979) (1807) ("[J]ust as lordship showed that its essential nature is the reverse of what it wants to be, so too servitude in its consummation will really turn into the opposite of what it immediately is.").

technology in the years to come. Yet, in one respect, the possibilities and dangers of robotics seem clearer to us in 2015 than the promise and threat of the Internet did in the early 1990s. Already we think we grasp what the law of robotics will be "about."

One reason why we may have strong notions of what robotics can do and the problems it can cause is that we have a rich literature of literary examples that concern the substitution effect. Indeed, the substitution effect may be one of the oldest tropes in artistic history. The legendary craftsman Daedalus, for example, was said to have created statues so lifelike that they had to be chained up to keep them from running away.[47] Throughout human history people have told stories of human artifacts that "come alive" and substitute for various functions of human beings. Examples range from Pygmalion's statue Galatea, to the Golem of Prague, to *Rossum's Universal Robots*, to the endless variations on the tale of Frankenstein, to Isaac Asimov's robot stories and Three Laws of Robotics, to familiar robotic characters in television shows and movies like *Lost in Space*, *Terminator*, *Star Wars*, and *Star Trek: The Next Generation*.

These literary and artistic sources—and many others we could name— already give us a sense of what robots might do and the problems that their introduction into society might cause. We should not, however, be too confident that we have seen this particular movie before, or confuse the vividness of our imagination with certainty about what technology and society have in store for us.

---

47. *Daedalus*, MYTHS ENCYCLOPEDIA, http://www.mythencyclopedia.com/Cr-Dr/Daedalus.html.