This article has been accepted for publication in IEEE Transactions on Technology and Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTS.2022.3233776



# Abstracted Power and Responsibility in Computer Science Ethics Education

Tina L. Peterson, Rodrigo Ferreira, and Moshe Y. Vardi

***Abstract*—As computing becomes more powerful and extends the reach of those who wield it, the imperative grows for computing professionals to make ethical decisions regarding the use of that power. We propose the concept of abstracted power to help computer science students understand how technology may distance them perceptually from consequences of their actions. Specifically, we identify technological intermediation and computational thinking as two factors in computer science that contribute to this distancing. To counter the abstraction of power, we argue for increased emotional engagement in computer science ethics education, to encourage students to feel as well as think regarding the potential impacts of their power on others. We suggest four concrete pedagogical approaches to enable this emotional engagement in computer science ethics curriculum, and we share highlights of student reactions to the material.**

***Index Terms*—power, abstraction, responsibility, social impact, emotional engagement, ethics**

## I. Introduction

In a darkly funny scene in the American TV series "The Good Place [1]," the Trolley Problem is made all too real for a moral philosophy professor who had previously encountered it only as a thought experiment. The professor, Chidi, is pathologically indecisive, and his teacher/tormentor Michael wants to force Chidi to make an actual decision by putting him in the driver's seat of a real (or virtually believable) trolley. From the look of horror on Chidi's face, it is clear that he would much rather consider the ethical dilemma from the safety of his armchair, comfortably removed from flesh-and-blood consequences.

As ethics educators in computer science, we are all too familiar with the discomfort that Chidi experiences. Our students often want to believe that ethical dilemmas can be resolved through reason alone. Like Chidi, they are confident they can think through a problem and arrive at an ethical solution. The teacher, Michael, knows that thinking is not enough; he makes Chidi *feel* the responsibility he has and the impact his decisions have on others. Only then can Chidi meaningfully wrangle with the problem.

The Trolley Problem cannot by itself enable such meaningful engagement. Because it is a thought experiment, it can only ever capture abstract power. It cannot communicate the messy, painful, disturbing ground truth of human decision-making, in which terror, injury, loss of life, unfair treatment, or trauma can result from the choice of track one versus track two.

Abstract power is familiar enough in our everyday lives; we make choices all the time, and we seldom witness first-hand the consequences of those choices on others. The 21st-century world is a complex place; abstract power seems inevitable in a global, interconnected society. What we are concerned with in this essay is how power can *be abstracted by technology*, and how that abstraction distances computing professionals from the impacts of their decisions and actions. Further, we are interested in how the emotional distance caused by abstracted power impacts computing professionals' capacity to reason ethically about how they should use their power.

In this paper we introduce the concept of abstracted power, a pedagogical tool that has proven useful in our classrooms, based on a combined 10-years' experience of teaching ethics in computer science departments. First, we define abstracted power and briefly explain how we conceptualize it with our students. Second, we review two significant challenges that technology–ethics educators face today: increasing intermediation in technological development, and the reason–centric nature of "computational thinking." Third, we describe four specific pedagogical approaches we use to add friction to this mode of computational thinking; our intent is to encourage our students to engage emotionally with technology and its impacts. Following each approach we summarize student reactions that illustrate their engagement with the material. We conclude by discussing our reflections on abstracted power as a pedagogical tool and acknowledging limitations on the observations we describe here.

## II. Introducing Abstracted Power

We define abstracted power as a human actor's influence or control over a system, process, or dataset which, as a function of the technology that enables it, obscures or distances the human actor from consequences of that influence or control. The emotional consequences of such decisions are obfuscated by a technological intermediary — a lever, joystick, keyboard,

Tina L. Peterson is in the Department of Computer Science at the University of Texas at Austin, Austin, TX 78712 (e-mail: peterson@cs.utexas.edu).

Rodrigo Ferreira is in the Department of Computer Science at Rice University, Houston, TX 77005 (e-mail: rf29@rice.edu).

Moshe Y. Vardi is in the Department of Computer Science at Rice University, Houston, TX 77005 (e-mail: vardi@cs.rice.edu).

This article has been accepted for publication in IEEE Transactions on Technology and Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTS.2022.3233776



or other user interface. The consequences still occur, but they have been so spatially and temporally removed that the human actor may find them easy to dismiss.

In our classes, we often introduce the concept of abstracted power with a modification of the Trolley Problem. In its most traditional form, the thought experiment posits a person (the agent) standing next to a lever that can alter the direction of a runaway trolley. The agent is given a choice: do nothing, and the trolley will kill five people who are standing in its path on one track, or pull the lever, which will change the trolley's course to a second track where it will only kill one person.

From a utilitarian perspective, pulling the lever seems like an easy choice. Doing nothing will kill five, while pulling the lever will kill only one. In purely numerical terms, the choice seems obvious. Replacing the lever with another human being, however, complicates the situation. In a common modification to the thought experiment, there is a person standing on a bridge overhanging the tracks. Rather than pulling a lever, the agent must decide whether to push this person onto the tracks. In the scenario, this person's fatal fall would stop the trolley and save the five people in its path. People given this modified situation tend to be reluctant to push the person, even though the resulting numbers of living and dead are the same as if they had pulled a lever.

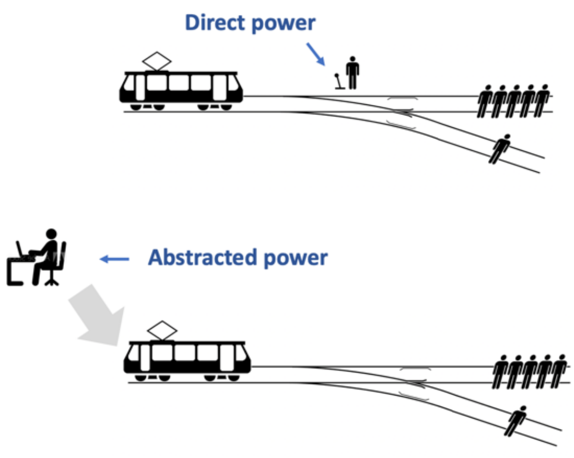



*Figure 1: Direct vs abstracted power in the Trolley Problem (Image adapted from McGeddon, CC BY-SA 4.0 via Wikimedia Commons)*

In a different modification of the Trolley Problem, we extend the 'lever' further by taking the agent out of the scenario altogether; we sit them at a computer terminal that is entirely removed from the action. In the diagram (Fig. 1), it is clear that the agent at the computer is still in control of the trolley via the code they are writing, but the agent is at a great enough distance that they never witness the consequences of that code's execution.

Students identify the differences in these scenarios immediately. They observe that the agent's experience of the situation depends on the distance between them and the scene. There is a difference between pushing a person and pushing a lever, and even more between pushing a lever and writing code far away from the scene of the action. Yes, they concede, the outcome is the same. But the distance from the impact makes all the difference. Some describe the horror in the man's face as they push him, facing his friends or family, etc. They contrast this with pushing the lever, which in many depictions of the trolley problem takes place from a safe distance, away from the judgment of those involved or of other bystanders, and away from having to perceive the consequences. They further contrast pulling the lever with writing code that, upon its execution at an unspecified point in the future, would determine the trolley's path and affect people they would never see. Their discussions of these scenarios demonstrate that the concept of abstracted power makes intuitive sense to them.

## III. The Challenges of Technological Intermediation and Computational Thinking

Having introduced the concept of abstracted power, we now turn to the challenges this pedagogical tool helps us address as ethics educators. In our classes we celebrate technology's capacity to improve people's lives, and we interrogate its other capacities. Who stands to benefit from a particular technology? Who stands to suffer under it? Whom might it empower, and whom might it oppress? And, most importantly for our purposes in this paper, how might the technology itself interfere with the felt responsibility of technologists regarding the impacts of the technology on others?

For decades, scholars and practitioners inside and outside of computer science have addressed the ethical questions that emerge with the rise of computational innovations and related social developments [2]–[7]. In the last few years new kinds of questions have emerged that renew focus on the significance of tech ethics and challenge it in new ways. Tech developers – and in particular the leaders of Big Tech corporations in Silicon Valley – have been criticized for lacking empathy and a sense of responsibility regarding the social problems that their technologies have created or exacerbated. Specifically, decisions made by technology corporations have negatively impacted society's capacity for civil dialogue, privacy, and fairness, among others [8]–[10].

Two factors make confronting this dearth of responsibility particularly difficult. The first is technological intermediation: technologies over time have changed to allow for a greater degree of opacity between tech developers and their users. In 1980, Langdon Winner observed, "Technological change expresses a panoply of human motives, not the least of which is the desire of some to have dominion over others [11]." Today technologists can exert influence over others much further removed from the loci of their actions and at an enormous scale. Billions of people across the world now use services and platforms developed by corporations such as Google and Meta (headquartered in Silicon Valley) that sell user data in exchange for advertisements, which aim at "gradually" and "imperceptibly" changing people's behavior thousands of miles away [12]–[14]. This is abstracted power writ large.

In other cases of technological intermediation the stakes are far higher. For example, much has been written about the distancing of military drone operators from the consequences of their actions. Chamayou [15] identifies drone technology as a "moral shock absorber" that enables operators to compartmentalize; "they kill during the day and go back home at night." Asaro [16] describes a combination of "remote agency" with surveillance and a particular division of labor that enables a "bureaucratization of killing."

Given the growing distance technological intermediation puts between tech's creators and its end users (and other human stakeholders), it can become increasingly difficult for developers – and those learning to develop these technologies – to understand the impact of what they build and put out into the world. Computer science students certainly use many of these technologies within their cultural milieu – but to what extent do they understand the impact that tech has on people they will never meet and cultures they will never encounter?

The second factor that interferes with the communication of responsibility is the nature of computational thinking. As formulated by Wing [17] and widely deployed in computer science education, "computational thinking" is all about abstraction. Computer science education has long focused on training students to think in terms of variables, data types, and algorithms rather than the meanings or the people behind them. CS students are not usually encouraged — unless by design in an ethics module — to connect a row in a dataset with the human being it represents. Differences among people become frequency distributions, and vulnerable users become edge cases. Such abstraction is necessary to the complex calculations that form the basis of computing.

The human messiness and unpredictability that are expected in the social sciences are antithetical to computer science, and may be scrubbed from a system or dataset to enable high–level reasoning about it. Though reason is certainly necessary for resolving all kinds of practical problems, psychologists agree that it is not often the most effective tool to handle situations involving the pain of others [18]–[20]. Computational thinking may even hinder understanding of the true social impacts of technology. It is easy to lose touch with on–the–ground emotional repercussions of one's work when one performs that work at a great, reason–centric elevation.

Abundant support exists for supplementing training in computational thinking with emotional engagement. Science has shown that emotions *are* a part of moral reasoning. Philosophers have long argued not only that they are, but that they *should* be. Here we elaborate on these two points.

From a scientific perspective, emotional engagement has been shown to be critical in ethical decision–making. Abundant research in neuroscience suggests that the region of the human brain associated with emotional processing activates when people engage in "moral cognition." Koenigs et al. [21] review research on fMRI imaging of the ventromedial prefrontal cortex (VMPC), the region of the brain "necessary for the normal generation of emotions and, in particular, social emotions." Evidence shows that the VMPC consistently engages when people view "morally salient pictures" and when they are asked to judge the rightness or wrongness of simple moral statements. Other studies described by Koenigs et al found that damage to the VMPC region of the brain was associated with impaired moral judgment. What this significant body of research suggests is that, at a neurochemical level, humans engage emotionally in moral judgments and moral decision–making, and that a deficit in emotional engagement tends to make a person less able to think in moral terms.

From a theoretical perspective, philosophers and literary writers have long argued for the ineluctable significance of feelings in moral decision–making. Rather than thinking of ethics as driven by abstract values or principles, such as achieving maximum utility or respect for another person's dignity, numerous philosophers going as far back as Aristotle have instead argued for ethics as a kind of "practical wisdom" – a habitual practice that embodies and encourages particular virtues, among them humility, compassion, and care [22].

Along similar lines, writers in areas such as phenomenology, feminism, cultural studies, and post–colonial theory have often emphasized the moral significance of feelings such as "care" [23] rather than abstract reason in moral decision-making, especially in consideration of historical social inequalities and of relations with asymmetrical power [24]–[26]. For this line of writers, the point in many ways has been to show not only that feelings are an intrinsic part of moral reasoning, but that they *should* be a part of it. As philosopher Martha Nussbaum [27] has put it, "Emotions aren't just mindless urges; they contain thoughts about matters of importance."

In this section we first examined two significant factors that contribute to the abstraction of power: the intermediation that widens the gap between a technology's developers and its end users, and the disciplinary tendency toward computational thinking in computer science. We then made a case for supplementing computational thinking with emotional engagement, which science and philosophy have shown to be both necessary and prudent to moral reasoning. Next we examine approaches we have used in our classrooms to encourage emotional engagement and to narrow the gap between developer and user that technology tends to widen.

## IV. Using Friction to Facilitate Emotional Engagement

Pedagogical approaches to encourage reckoning with the social impacts of technology abound. At an academic level, numerous books today suggest approaching technological design from an increasingly interdisciplinary perspective. This would include psychology, sociology, and philosophy, as has long been the case, but also critical theory, cultural studies, disability studies, gender studies, and other areas that can help illuminate the historical impact of technology on social minorities, and draw attention to historical biases in the social construction of technology. For example, in Data Feminism D'Ignazio and Klein [28] argue readers should move past simple "data ethics" and concern for topics such as bias and algorithms, and instead shift their focus to the more complex concept of "data justice." This includes focusing on the particular histories of oppression that have helped to create

This article has been accepted for publication in IEEE Transactions on Technology and Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTS.2022.3233776



those biases that are then instantiated in algorithms. Vallor [29] has argued for a shift from traditional frameworks in normative philosophy, such as utilitarianism and deontology, which are based on abstraction, to the more flexible and feeling-oriented paradigm of virtue ethics – including feelings such as care, compassion, and empathy – when dealing with contemporary technologies.

Specifically for technology educators, Patrick [30] suggests using the tools of Science and Technology Studies (STS) to help engineers in the classroom to consider the specific populations who will use a technology, along with its broader impact on society. Other scholars have suggested curricula incorporating role-playing [31] and fiction [32], which can also help students integrate more of an affective element in their moral considerations. Along similar lines, two of the authors have addressed the value of bringing a greater focus on social justice [33] and care [34] into technology and ethics classrooms.

We perceive a thread common in many of these approaches: they add *friction* to students' experiences of technology. Friction encourages deliberation, mindfulness, and critical reflection. Here we identify a point of tension: an underlying goal of much computer science is to remove friction and improve efficiency; to design systems that operate more smoothly and predictably, or that require less energy or action on the part of the user. Struggles and delays and user effort are usually considered barriers to be removed, to optimize user experience or to make a process work faster [35]. Friction is the enemy of speed and efficiency. However, as argued by Vardi [36], removing friction can result in more fragile and less resilient systems. For example, the minimization of time delay in algorithmic securities trading has made Wall Street more efficient, but less resilient [37]. Tomalin [38] outlines a typology of different kinds of online friction, and argues that those which protect the user are desirable and should not be eliminated.

For our purposes – to encourage emotional engagement in a classroom – some friction is essential. In its most basic form in a user interface, friction makes the human actor pause and double-check an action they're about to take, and it is crucial in a safety-critical system. It is the "are you sure?" prompt you see before deleting something, the dual-factor authentication required for secure logins, and the flip-top cover over the big red button in a control panel. Friction can also be meaningful to the developer of technology: it can slow them down and force them to think about the potential impacts of the design decisions they make and the code they write. It can add back in the details that were abstracted away by computational thinking. Friction can reveal the human messiness and power relations embedded in a system or dataset; instead of feeling comfort and confidence, the developer may experience uncertainty and a greater sense of felt responsibility.

Friction can bring the developer working at a great elevation back down to the ground truth; it encourages a reckoning with who may be helped and who may be harmed by what they are building. In our computer science ethics courses, we try to add friction to students' thought processes to encourage them to transcend computational thinking and engage emotionally. Next we describe four approaches that have proven effective at introducing friction and encouraging students not just to think, but to feel. Following each, we share highlights of students' reactions to them.

## V. Four Pedagogical Approaches

### A. Write ten lines of IF–THEN–ELSE code

The first approach adds emotional friction to an everyday task of a computing professional: writing code with IF–THEN–ELSE logic. We assign students to write 10 lines of code to handle the collision-avoidance policy of an autonomous vehicle. In the prompt, we establish that there is insufficient time and space to apply the brakes and avoid the collision, so the vehicle must evaluate inputs and make a decision. It is the Trolley Problem, now in the hands of (potential future) autonomous-vehicle developers. The students' code dictates who or what should be crashed into, and who or what should be saved. The students write the code in small groups, publish their work on the course's Canvas, and then discuss the experience as a class. When given this assignment, many students squirm at the implied responsibility in their roles. Their reactions to this exercise — feeling uncomfortable "playing God," and at a loss regarding how to make the "right" decision — demonstrate their emotional engagement.

The contents of their code are also revelatory. Every semester, several groups reduce the task to a utilitarian equation, the oversimplified essence of the Trolley Problem: their code compares the numbers of people potentially affected by either clause in their IF–THEN–ELSE logic and takes the action that would cost the fewest lives. Others propose measuring the ages of potential victims and maximizing for years–left–to–live. A handful of students always include an ELSE clause with a random selection function to determine the path of the vehicle. When asked to explain their code, they shrug sheepishly and explain that they're more comfortable leaving it up to chance than making the decision deliberately.

### B. Unplug and reflect

The second pedagogical approach makes the effects of technology personal and encourages students to reflect on the user's perspective. In an "unplug" assignment, we ask students to temporarily abstain from using specific apps or devices that they find addictive [39], [40]. In class, the students collectively create a list of all the apps they consider potentially addictive; this crowd-sourced list becomes the basis for the unplug. We ask them to write a short paper in which they reflect on the experience. Some of us assign readings such as Odell [41] to help guide students' thinking, and others ask the students to reflect on particular concepts such as attention and boredom.

In their written reflections on the experience, the engagement–promoting features of these apps (the infinite scroll, for example) are thrown into sharp focus. Reacting to the ultimate form of friction — an all–out block on their usage of a particular app or device — students realize how deeply affected they are by its features. They frequently describe feeling angry and powerless against the pull of their notifications, and having to deliberately stop their fingers from idly responding to nudging requests absent their conscious will. The power of these features over them causes them anxiety and stress, and

This article has been accepted for publication in IEEE Transactions on Technology and Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTS.2022.3233776



many acknowledge that the apps' developers — a group they may join one day — have a responsibility for the apps' addictive design.

### *C. Read and discuss case studies*

Reading and discussing journalistic case studies is the third pedagogical approach we use; non–fiction narratives introduce friction into students' understanding of technology that they have studied in technical classes and may take for granted. High–quality tech journalism illuminates power relations among tech's developers and its end–users and other stakeholders. One example is ProPublica's "Machine Bias" [42], an engaging account of people impacted by the racial bias trained into a recidivism–risk prediction algorithm. Another is a report in The New York Times [43] about an African–American father of two who was arrested in front of his wife and children after being wrongly identified by a facial recognition system in connection with a robbery. A compelling story by Hao and Freischlad [44] takes readers into the everyday lives of gig workers in Indonesia, who find ways to push back against the ride-hailing algorithm that dictates the terms of their livelihoods. Such journalistic case studies often prompt strong emotional reactions from students. These stories reveal the names, faces, and life stories behind rows of data, and few students feel unmoved by accounts of people treated unfairly, lives lost, or opportunities given or taken away by lines of code.

### *D. Critically reflect on relevant science fiction*

A fourth and very popular pedagogical approach that encourages emotional engagement is to assign students to read or view science fiction and write about it critically. Reading fiction that is emotionally transporting has been shown to increase the reader's empathy [45], and science fiction enables students to identify with characters impacted by imagined or emerging technology. Black Mirror [46] is a television series of stand-alone episodes, each of which imagines the social or cultural impact of a particular futuristic technology. The episodes "Nosedive," "Men Against Fire," and "15 Million Merits" are especially thought–provoking. The film Minority Report [47] imagines dystopic consequences of applying predictive data analytics to criminal justice. The film Ex Machina [48] examines the moral status of artificial intelligences as well as social relations between humans and robots. Among books, The Circle [49] stands out for its critique of social media's leveraging of its users' and employees' private information. The Ministry for the Future [50] encourages students to think through environmental consequences of technology as well as social conflicts and inequalities exacerbated by climate change. Thought–provoking science fiction abounds in the 21st century; these are just a few examples that can promote emotional engagement and critical reflection in the CS ethics curriculum.

Similar to their reactions to journalistic case studies, students' emotional reactions to works of fiction about technology are quite strong. Black Mirror episodes have proven especially thought–provoking and affecting. Every semester, several students identify in their analysis papers disturbing parallels between the technology depicted in the episode "Men Against Fire" and propaganda used by Nazis. One student wrote that he felt sick and heartbroken at the plight of the main characters in "15 Million Merits."

In the previous section we established a theoretical basis for adding friction to computer-science curriculum in order to facilitate emotional engagement and encourage a sense of felt responsibility. We then introduced four specific approaches we have used in our own classrooms to accomplish this, and summarized students' reactions to them.

## VI. Observations and Discussion

In this section we share general observations of students' responses to the approaches described above and to the concept of abstracted power. We do not mean to represent this work as the results of an experiment or a systematic testing of a hypothesis. We have used these discussion prompts, readings, activities, and assignments in uncontrolled settings with diverse groups of computer science students at two institutions – private and public – and we share our observations with the hope that our fellow educators may find something of use here.

It is apparent from many students' final papers that the concept of abstracted power continues to resonate with them throughout the semester. Many talk about the importance of making computing professionals more aware of their power; they connect this with the capacity to make more socially responsible decisions. Some specifically address the distance between technology developers and their users as a gap that needs to be bridged.

Over the past several years in our computer science ethics classes, we have come to believe that abstracted power is an important and useful concept. It resonates with students and helps explain much of what is going on in the world right now. Students 'see it' in many different scenarios and case studies throughout the class. They identify it in the behavior and decisions of tech leaders, and even in the behavior of themselves and their classmates.

The concept also serves us as educators; it helps us communicate to students the different means of technological intermediation, as well as the tendencies and potential blind spots of computational thinking. Its conceptual inverse – that is, the de-abstraction or concretization of power – helps us and our students think through and feel the responsibilities we have as creators of technology.

In some of our classrooms, we begin the semester with a variation on the motto originally attributed to the sword of Damocles and popularized by Spider-Man: "With abstract power comes the need for more concrete responsibility." This simple change to the original motto (which is familiar to most of the students) helps us convey the fact that, though technology today provides more power than ever, and that more abstractly, this does not excuse them from moral responsibility for their technology-enabled actions.

## VII. Conclusion

In this article we identified two challenges to ethics education in computer science: increasing technological intermediation and the disciplinary primacy of reason over feelings in computational thinking. We discussed several strategies by which scholars and educators are attempting to address these challenges, encouraging more critical perspectives and pushing the visibility of users to the fore.

Then we described how, following the spirit of some of these strategies, we have taken on this challenge in our classroom by applying the concept of "abstracted power." As evidenced by students' responses, this concept has helped many of them wrangle with the distance they may feel from consequences of their future work. It also facilitates emotional engagement and a sense of felt responsibility for the impact they may have on others. Ultimately, we hope this will help students become more responsible computing professionals, insofar as they can make sound ethical choices attuned to their own feelings and those of others.

We described here the success we have had in achieving certain learning objectives in our classrooms, but we also acknowledge the limitations of our work. We have not systematically quantified our students' attitudes or reactions nor measured them over time. For this reason, we acknowledge that our work is not decisive in terms of quantitative evidence. We also recognize that our pool of students may not be generalizable to a larger population.

Despite these limitations, we hope that the work presented here can continue to encourage reflection on the problem of the lack of empathy among developers of technology, and can help inspire further pedagogical innovation addressing this problem in computer science education and related fields. Ethics education, along with any endeavor that aims to influence human attitudes and behaviors, is an imperfect work–in–progress that benefits from collaboration and feedback. With this article, we hope to introduce a concept that can meaningfully resonate with computer science students and which, we hope, would lead to fruitful conversations with our fellow educators.

## References


[1] M. Schur et al., "The good place," 2016.
[2] J.H. Moor, "What Is Computer Ethics?" *Metaphilosophy,* vol.16, no. 4, pp. 266–75, 1985.
[3] K. Miller, "Integrating computer ethics into the computer science curriculum." *Computer Science Education*, vol. 1, no. 1, pp. 37-52, 1988.
[4] C.D. Martin and E.Y. Weltz, "From awareness to action: Integrating ethics and social responsibility into the computer science curriculum." *ACM Sigcas Computers and Society*, vol. 29, no. 2, pp. 6-14, 1999.
[5] T.W. Bynum, Computer ethics: Its birth and its future. *Ethics and Information Technology*, vol. 3, no. 2, pp. 109–112, 2001.
[6] M.J. Quinn, "On teaching computer ethics within a computer science department." *Science and Engineering Ethics*, vol. 12, no. 2, pp. 335-343, 2006.
[7] B.J. Grosz, D.G. Grant, K. Vredenburgh, J. Behrends, L. Hu, A. Simmons, and J. Waldo. "Embedded EthiCS: ntegrating ethics across CS education." *Communications of the ACM*, vol. 62, no. 8, pp. 54-61, 2019.
[8] B. Olaniran and I. Williams,"Social media effects: Hijacking democracy and civility in civic engagement," in *Platforms, Protests, and the Challenge of Networked Democracy*. Springer, 2020, pp. 77–94.
[9] J. P. Bagrow, X. Liu, and L. Mitchell, "Information flow reveals prediction limits in online social activity," *Nature human behaviour*, vol. 3, no. 2, pp. 122–128, 2019.
[10] V. Eubanks, *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.
[11] L. Winner, "Do artifacts have politics?" *Daedalus*, vol. 109, no. 1, pp.121-136, 1980.
[12] C. Fuchs, *Social Media: A Critical Introduction* (2nd edition). SAGE Publications Ltd., 2017.
[13] J. Lanier, *Ten Arguments for Deleting Your Social Media Accounts Right Now*. Henry Holt and Co., 2018.
[14] S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (1st edition). PublicAffairs, 2019.
[15] A. Schwartzbrod, "Grégoire Chamayou – War is Becoming a Telecommuting Job for Office Workers (interview)." Tripleampersand.org, 2015.
[16] P.M. Asaro, "The labor of surveillance and bureaucratized killing: new subjectivities of military drone operators." *Social Semiotics*, vol. 23, no. 2, pp. 196-224, 2013.
[17] J. Wing, "Computational Thinking." *Communications of the ACM*, vol. 49, no. 3, pp. 33–35, 2006.
[18] T. Wiseman, "A concept analysis of empathy." *Journal of Advanced Nursing*, vol. 23, no. 6, pp. 1162–1167, 1996.
[19] M.L. Hoffman, *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge University Press, 2000.
[20] S. Turkle, *Reclaiming Conversation: The Power of Talk in a Digital Age*. Penguin Press, 2015.
[21] M. Koenigs, L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. Damasio, "Damage to the prefrontal cortex increases utilitarian moral judgements," *Nature*, vol. 446, no. 7138, pp. 908–911, 2007.
[22] Aristotle, *Aristotle's Nicomachean Ethics* (R. C. Bartlett & S. D. Collins, Trans.) University of Chicago Press, 2012.
[23] C. Gilligan, *Joining the resistance*. John Wiley & Sons, 2011.
[24] P.T. Clough and J. Halley (eds.), *The Affective Turn: Theorizing the Social*. Duke University Press, 2007.
[25] M. Gregg and G.J. Seigworth, (eds.), *The Affect Theory Reader*. Duke University Press, 2010.
[26] S. Ahmed, *The Cultural Politics of Emotion*. Routledge, 2014.
[27] M. Nussbaum, "Discussing Disgust" (J. Sanchez, interviewer), Reason.com, 15 July, 2004.
[28] C. D'Ignazio and L.F. Klein, *Data Feminism*. The MIT Press, 2020.
[29] S. Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press, 2018.
[30] A. Y. Patrick, "Bringing Care and Concern to Engineering Students Through STS Knowledge." *IEEE Transactions on Technology and Society*, vol. 2, no. 2, pp. 103–104, 2021.
[31] S.A. Doore, C. Fiesler, M.S. Kirkpatrick, E. Peck, and M. Sahami, "Assignments that Blend Ethics and Technology." *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pp. 475–476, 2020.
[32] E. Burton, J. Goldsmith, and N. Mattei, "How to Teach Computer Ethics through Science Fiction." *Communications of the ACM*, vol. 61, no. 8, pp. 54-64, 2018.
[33] R. Ferreira and M.Y. Vardi, "Deep Tech Ethics: An Approach to Teaching Social Justice in Computer Science." *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pp. 1041–1047, 2021.
[34] R. Ferreira and M.Y. Vardi, Computer Ethics and Care: An Activity for Practicing "Deep" Attention. *Teaching Ethics*, vol. 20, no. 1/2, pp. 139–156, 2020.
[35] M.Y. Vardi, "Fricative computing," *Communications of the ACM*, vol. 56, no. 5, pp. 5–5, 2013.
[36] M.Y. Vardi, "Efficiency vs. Resilience: Lessons from COVID-19." In H. Werthner, E. Prem, E.A. Lee, and C. Ghezzi, (eds) *Perspectives on Digital Humanism*. Springer, 2022.
[37] E. Budish, P. Cramton and J. Shim, "The high-frequency trading arms race: Frequent batch auctions as a market design response." *Quarterly Journal of Economics*, vol.130, no. 4, pp. 1547-1621, 2015.
[38] M. Tomalin, "Rethinking online friction in the information society." *Journal of Information Technology*, 2022.
[39] T. L. Peterson, "Syllabus: Ethics and accountability in computer science," 2019.
[40] R. Ferreira and M. Y. Vardi, "Computer ethics and care: An activity for practicing "deep" attention," *Teaching Ethics*, vol. 20, no. 1/2, pp. 139–156, 2020.

This article has been accepted for publication in IEEE Transactions on Technology and Society. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTS.2022.3233776



[41] J. Odell, *How to Do Nothing: Resisting the Attention Economy*. Melville House, 2019.
[42] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, “Machine bias.” ProPublica, 2016.
[43] K. Hill, “Wrongfully accused by an algorithm,” *The New York Times*, 2020.
[44] K. Hao and N. Freischlad, “The gig workers fighting back against the algorithms.” *MIT Technology Review*, 2022.
[45] P. M. Bal and M. Veltkamp, “How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation,” PloS one, vol. 8, no. 1, p. e55341, 2013.
[46] C. Brooker et al., “Black mirror,” 2012.
[47] S. Spielberg, dir., “Minority report,” 2002.
[48] A. Garland, dir., “Ex machina,” 2014.
[49] D. Eggers, *The Circle*. Knopf, 2013.
[50] K. S. Robinson, *The Ministry for the Future: A Novel*. Orbit, 2020.

**Tina L. Peterson** received the B.Sc. degree in Journalism from the University of Colorado at Boulder, in 2000, the M.A. degree in Critical Theory and Cultural Studies from the University of Nottingham, in 2003, and the Ph.D. degree in Mass Media and Communication from Temple University, in 2012.

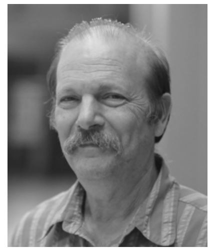
She is an Assistant Professor of Instruction in Computer Science at the University of Texas at Austin, where she teaches undergraduate and graduate courses on ethics and social responsibility in computer science, including AI and robotics. She is senior personnel on a National Science Foundation Research Traineeship program, The Convergent, Responsible, and Ethical AI Training Experience (CREATE) for Roboticists.

In addition to teaching and research, she is the author of the children’s book *Oscar and the Amazing Gravity Repellent* (Capstone, 2015).

**Rodrigo Ferreira** has a B.A. degree in Philosophy (with honors) and Psychology (2009), a M.A. in Humanities and Social Thought (2014), and Ph.D. in Media, Culture, and Communication (2019), all from New York University. He is currently an Assistant Teaching Professor in Computer Science at Rice University, where he is responsible for teaching all ethics courses and developing ethics-related curricula in Computer Science. In this position, Rodrigo currently teaches undergraduate and graduate courses on the ethics of Computer Science, Data Science, and AI and Robotics.

In collaboration with Dr. Moshe Vardi, Rodrigo has also developed “Deep Tech Ethics” as a pedagogical approach that seeks to orient computer science education toward greater focus on affective care, historical power inequalities, and social justice. His work on this project has been published in Teaching Ethics and in the Proceedings for the Association for Computing Specialized Interest Group in Computer Education and been presented at conferences and events in the United States, Latin America, and Europe. From 2019 to 2021, Rodrigo was a Postdoctoral Researcher in Technology, Culture, and Society with the Rice Academy of Fellows and the Department of Computer Science at Rice University.

In addition to his pedagogical practice and research, Rodrigo also translated late Mexican-Ecuadorian philosopher Bolívar Echeverría’s Modernity and “Whiteness” to English (Polity 2019) and co-authored a public policy recommendation report on “AI Ethics” as part of the Mexican National AI Agenda 2020-2030.



**Moshe Y. Vardi** is University Professor and the George Distinguished Service Professor in Computational Engineering at Rice University. He is the co-recipient of three IBM Outstanding Innovation Awards, the ACM SIGACT Goedel Prize, the ACM Kanellakis Award, the ACM SIGMOD Codd Award, the Blaise Pascal Medal, and the IEEE Computer Society Goode Award. He is the author and co-author of over 700 papers, as well as two books: "Reasoning about Knowledge" and "Finite Model Theory and Its Applications".

He is a Guggenheim Fellow, as well as Fellow of the American Mathematical Society, the Association for the Advancement of Artificial Intelligence, the Association for Computing Machinery, the American Association for the Advancement of Science, the Institute for Electrical and Electronic Engineers, and the Society for Industrial and Applied Mathematics. He is a member of the US National Academies of Science and of Engineering, the American Academy of Arts and Science, the European Academy of Science, and Academia Europaea.

He holds honorary doctorates from the Saarland University in Germany, Orleans University in France, UFRGS in Braziland, University of Liege in Belgium, the Technical University of Vienna, the University of Edinburgh, the University of Grenoble, and the University of Gothenburg. He is Senior Editor of the Communications of the ACM, after having served for a decade as Editor-in-Chief. Vardi's interests focus on automated reasoning, a branch of artificial intelligence with broad applications in computer science, including database theory, computational-complexity theory, multi-agent systems, computer-aided verification, and teaching logic across the curriculum.