

Using “The Machine Stops” for Teaching Ethics in Artificial Intelligence and Computer Science

Emanuelle Burton
University of Chicago
Chicago, Illinois, USA
eburton@uchicago.edu

Judy Goldsmith
University of Kentucky
Lexington, Kentucky, USA
goldsmi@cs.uky.edu

Nicholas Mattei
Data61 and UNSW
Sydney, Australia
nicholas.mattei@nicta.com.au

Abstract

A key front for ethical questions in artificial intelligence, and computer science more generally, is teaching students how to engage with the questions they *will* face in their professional careers based on the tools and technologies *we* teach them. In past work (and current teaching) we have advocated for the use of science fiction as an appropriate tool which enables AI researchers to engage students and the public on the current state and potential impacts of AI. We present teaching suggestions for E.M. Forster’s 1909 story, “The Machine Stops,” to teach topics in computer ethics. In particular, we use the story to examine ethical issues related to being constantly available for remote contact, physically isolated, and dependent on a machine — all without mentioning computer games or other media to which students have strong emotional associations. We give a high-level view of common ethical theories and indicate how they inform the questions raised by the story and afford a structure for thinking about how to address them.

Introduction

Work as an AI or any technology development profession often requires engagement with ethical issues on a regular basis. It is incumbent on computer science professionals to educate both students and the general public about the ethical issues that arise around the creation and use of AI and technology. The widespread adoption of AI and related technologies put the creations and the creators in situations that call not only for technical decisions but for ethical ones. The precise criteria of moral judgment varies according to different schools of ethical theory, but most readers can easily recognize that the powers at stake in these new technologies can be used for good or for ill. There are different ways to ask ethical questions, as well as different ways to answer them, and as developments in technology offer new frontiers of possibility, asking and answering basic ethical questions becomes an unavoidable dimension of our work in computer science (Rogaway 2015).

Our position is that computer science professionals have a responsibility to train students to recognize the larger ethical issues and responsibilities that their work as technologists may encounter. To this end we have, in the last several years, published work on our course “Science Fiction

and Computer Ethics” (Mihail, Rubin, and Goldsmith 2014; Burton, Goldsmith, and Mattei 2015). This course has been popular with students, as has our previous work running an undergraduate computer science course that uses science fiction to engage students about research (Goldsmith and Mattei 2011; 2014). We are not alone in this endeavor or this opinion as others have cited using science fiction as a gateway as it, “[...] often removes the intellectual and emotional resistance some students might at first feel towards the subject of ethics (Pease 2009).” Pease reports that using science fiction to teach *practical ethics* engages students from a variety of majors including computer science and engineering. Many courses with ethical considerations at their core have been offered in the past, examples include at Humboldt University at Berlin¹ and also a version focused on legal issues at Stanford². Courses in other fields use science fiction or literature for motivation or as a core part of a non-majors course (Bates 2011; Dils 1987). Scholars in other humanistic disciplines such as history and philosophy have also argued that literature is an invaluable teaching tool for ethics, among other topics (e.g., (Garcia Iommi 2011; Goering 2014; Rodwell 2013)).

ABET, the one of the largest accreditors of engineering and technology programs, requires instruction on professional ethics; we argue that we must go further. In computer science, as with many high consensus fields, there is a tendency to teach from authority and not encourage discussion and dissent (Colbeck 1998). This often leaves students with a didactic view of “truth” which leaves them unequipped to reason about situations which involve multiple correct answers, or to engage in ethical tradeoffs (Haworth and Conrad 1995; Perry 1980). These skills are important not only for engaging with ethical decisions but also for analyzing tradeoffs within their professional careers. While educating students in these regards is often left to the cross-disciplinary portions of university curricula (in the US) (Davis 2009) we argue that spending time focused on how these issues apply to students’ professional careers as technology developers is important and necessary. Indeed, using reading and writing to gain understanding and expand communication skills is important and there have been recent arguments to move

¹<http://waste.informatik.hu-berlin.de/Lehre/ws0910/dystopien/>

²<http://web.stanford.edu/class/cs122/>

these functions closer to the topics of core study for a student, rather than leave them to another department (Walker 1998; Hoffman, Dansdill, and Herscovici 2006; Bean 2011; McLeod and Soven 1992). Indeed, Rogaway argues that the lack of engagement by faculties in the political and ethical ramifications of computer science is a failing of computer science as an *academic field* (Rogaway 2015).

How To Think

Among researchers in the AI community, there are not only multiple sets of values, but different approaches to the theory of value, i.e., what we have a responsibility to do. We do not want to unify the field around a particular value system. Our work as computer scientists does not take place in a vacuum; whatever principles we establish for AI will have ethical and societal implications. As leaders in the field, we have a responsibility to lead the charge and drive the discussion about the impacts of our work. Indeed, Boyer argues that academics have a responsibility to engage students and the public with their research (Boyer 1997; 1996), and we have seen leadership in this area in recent years (e.g., the recent letter presented at IJCAI 2015 on autonomous weapons research³). We should continue to differ, as a community, about the appropriate value systems on which to build AI systems. This debate is healthy and allows us to form and revise our beliefs. The goal of teaching ethics is not to impose a value system on our students, but to inform them about the multiplicity of value systems while, at the same time, making them aware of the social ramifications of their work; that research, development, and implementation can be carried out in a variety of ways. We want to enable our students (and the broader society, if possible) to make informed, thoughtful, and ethical choices.

It is important to consider what it means for us to say that we want to inform our students *how* to think instead of *what* to think. For individuals with a technical, rational bent (i.e., computer scientists and engineers) it is tempting to assume that we can formulate a set of rules, agree on them, and be done. For people used to working with concrete, testable things (i.e., program languages or contained systems) — where the cause and effect that ties language to world is reliable, consistent, and not dependent on the interpretive lenses of the individual — universal laws for ethical reasoning make a kind of sense. This approach works when it comes to programming the robots against concrete actions, e.g., do not shoot humans.

When the ethical questions are more fuzzy, and require a framework for analysis, the issue becomes harder. There is a standard ethical dilemma, usually framed in terms of a brakeless trolley train hurtling toward individuals or groups of people on the tracks: do you throw a switch or throw a sufficiently large person onto the tracks so as to kill a smaller number of individuals, or do you refuse to be the agent of sacrifice? This becomes a real and instantaneous decision when we are driving, and must choose between bad options for a crash. What if the vehicle that crashes is a self-driving car? We are seeing headlines such

³http://futureoflife.org/AI/open_letter_autonomous_weapons

as “Why Self-Driving Cars Must Be Programmed to Kill” in MIT Technology Review (Why 2015), based on Bonnefon’s article, “Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?” (Bonnefon, Shariff, and Rahwan 2015). See also (Goodall 2014a; 2014b; Cummings and Ryan 2014; D’Onfro 2015). The central issue is that one cannot program the *roboticist* (person) with a universal system of rules in the same way one programs the robot itself; human languages do not always consider effects in the same way, in the world or even in the individual mind.

The very idea of a universally-applicable ethical doctrine has serious problems. A good example of the pitfalls implicit in such a project is the United Nations’ Universal Declaration of Human Rights (UDHR). The UDHR was commissioned in 1946, when an international commission was formed, and published two years later; but only after the American Anthropological Association (AAA) had withdrawn from the Commission, on the grounds that any attempt to codify a universal definition of the “right” way to be human cannot, by definition, take account of the particular social and ethical context of individual cultures, and that the cultures that had historically been most oppressed would be the most likely to be ignored or de-legitimized by any “universal” declaration. As Melville Herskovits wrote in a statement to the Commission in behalf of the AAA:

How can the proposed statement of rights be applicable to all human beings, and not be a statement of rights conceived only in terms of the values prevalent in countries of the Western Europe and America?... It will not be convincing to the Indonesian, the African, the Indian, the Chinese, if it lies on the same plane as like documents [such as the Declaration of Independence] of an earlier period. The rights of Man in the Twentieth Century cannot be circumscribed by the standards of any single culture, or circumscribed by the aspirations of any single people. Such a document will lead to frustration, not realization of the personalities of vast numbers of human beings (The Executive Board of the American Anthropological Association 1947).

Although the statement was not universally endorsed by anthropologists at its initial publication and has continued to provoke discussion in the anthropological community, these debates are about how to strike a balance between (on the one hand) the ability to make moral assessments, and (on the other) a sensitivity to cultural difference, not about whether the basic critique of the UDHR was valid. A helpful and lucid account of these continuing debates can be found in, e.g., Merry (2003). As these critics of the UDHR point out, the attempt to secure universal freedom through a universal set of criteria works, paradoxically, to codify inequality. Nor, as more recent work has shown, can this problem be reduced to a simplistic east-west culture clash; as Joanne Bauer writes, “even within the West, particularly in the United States, there are significant numbers of people who hold ideas of human rights that are in tension with the dominant liberal interpretation of international human rights (Bauer 2003).”

Although the precise status and possibilities of human rights discourse is still debated, scholars in both ethics and

anthropology agree that there is no way to formulate universal precepts of this kind that *do not* reaffirm, on some level, the kinds of social inequality they are designed to answer. The idea that a single system of laws or duties (deontological ethics) would solve all problems, and that our responsibility as teachers is to transmit those laws to students, eliminates the individual understanding and reasoning that we should be teaching our students how to do.

Science Fiction and Teaching

Stories — literature, plays, poetry, and other forms of narrative — have always been a way of talking about our own world, telling us what it's like and what impact our choices will have. Whether they are transmitted in print or through other media, stories play a potent role in shaping the thoughts and ideas of individuals, and the cultural norms of the societies in which they live. Indeed, since Socrates banned the poets from his ideal city of Kallipolis in Book X of Plato's Republic, on account of the dangers their work posed, philosophers and religious thinkers have gloried in and despaired of the power of literature to make or break a reader's convictions about the world. Many ethical scholars have even argued that literature is superior to philosophy in its ability to represent and address the ethical conundra of human experience. Martha Nussbaum, one of the preeminent exponents of this position, writes,

Reading [fiction] frequently places us in a position that is both like and unlike the position we occupy in life: like, in that we are emotionally involved with the characters, active with them, and aware of our incompleteness; unlike, in that we are free of the sources of distortion that frequently impede our real-life deliberations. (Nussbaum 1990)

We take science fiction in its broadest sense⁴, as the fantastical worlds or even the futuristic technology gives us a starting platform for discussion. The category of science fiction was first described by Hugo Gernsback, for whom the prestigious Hugo Prize is named, in the editorial to the first issue of *Amazing Stories* in 1926 as:

By 'scientifiction' I mean the Jules Verne, H G Wells and Edgar Allan Poe type of story — a charming romance intermingled with scientific fact and prophetic vision ... Not only do these amazing tales make tremendously interesting reading ? they are always instructive. ... New adventures pictured for us in the scientifiction of today are not at all impossible of realization tomorrow.

From this broad definition, almost any fiction dealing with sufficiently advanced technology *is* science fiction. Though the majority of the literary and philosophical establishment does not science fiction seriously as a venue for ethical thinking, this fact reflects longstanding biases in the field rather than the merits or possibilities of science fiction itself. Burton (2014) gives an in-depth treatment of the ingrained biases against science fiction and other forms of non-realist literature.

⁴Though the precise definition of Science Fiction is a matter of some debate, within the field, at the moment (Wallace 2015).

Using *any* fiction allows us to reframe recognizable human situations and problems in terms of unfamiliar settings and technology. Hence, any fiction, and especially science fiction in the case of technology, can be an ideal medium for raising and exploring ethical concerns. By presenting a familiar problem (such as conflicts between different social groups or the invasion of privacy) in unfamiliar terms and settings, a work of science fiction can limit a reader's ability to identify transparently with any one aspect or position. Indeed, Nussbaum states,

Since the story is not ours, we do not get caught up in the vulgar heat of our personal jealousies or angers or the sometimes blinding violence of our loves. (Nussbaum 1990)

Hence, we advocate science fiction for several reasons:

1. The use of futuristic or alien settings allows students to detach from political preconceptions and experience the dilemmas of plot and characters as something fresh.
2. It has so far proved popular with the students.⁵ They have perceived that the course would be a chance to get credit for something they enjoy but have not found time to do while in college/graduate school: read and watch science fiction. Their only complaint (on one individual's teaching evaluation) was that they were not provided with answers to the ethical dilemmas presented.
3. By its nature, science fiction promotes discussion of possible future technology, with a focus on social implications of that technology.⁶
4. Some of the science fiction chosen also posits new science infrastructure, and allows students to think about doing research and development outside of the fairly rigid industrial and academic boxes, driven by something other than current funding paradigms.

Teaching Guide: "The Machine Stops"

We illustrate this approach using E.M. Forster's 1909 story, "The Machine Stops." In the next section, we summarize the story. We then present some of the ethical issues that this story raises, and discuss how they can be addressed through the frameworks of utilitarian ethics, deontological ethics, and virtue ethics.

The Story

Imagine, if you can, a small room, hexagonal in shape, like the cell of a bee. It is lighted neither by window nor by lamp, yet it is filled with a soft radiance. There are no apertures for ventilation, yet the air is fresh. There are no musical instruments, and yet, at the

⁵The instructor, Goldsmith, received three teaching awards the first time it was taught, and consistently received excellent teaching evaluations for it.

⁶To quote an anonymous reviewer, "... you can discuss novel technological conundrums that may present interesting and new ethical questions."

moment that my meditation opens, this room is throbbing with melodious sounds. An armchair is in the centre, by its side a reading-desk that is all the furniture. And in the armchair there sits a swaddled lump of flesh — a woman, about five feet high, with a face as white as a fungus. It is to her that the little room belongs.

Thus begins E.M. Forster's 1909 story, "The Machine Stops". In it, this "swaddled lump of flesh," named Vashti, experiences the end of the world as she knows it, the world of the Machine. She lives a life identical, she believes, to all other people, in a room identical to all others. Her environment is mediated by the Machine. She controls it by stops and buttons, like an organ. She can call forth food, can cause a bed to be produced and the light darkened so she can sleep. She can communicate with the thousand or so people she knows via speaking tubes, and see them, albeit imperfectly, through the Machine. She can listen to music or to lectures, always in search of new ideas.

As the story starts, her son Kuno calls her from the other side of the world and says he wants to see her. She tells him, "I dislike seeing the horrible brown earth, and the sea, and the stars when it is dark. I get no ideas in an airship."

Vashti's knowledge of her world comes from the Book of the Machine, which is published by the Central Committee. It tells her what buttons to press to summon what she needs.

Sitting up in the bed, she took it reverently in her hands. She glanced round the glowing room as if some one might be watching her. Then, half ashamed, half joyful, she murmured 'O Machine! O Machine!' and raised the volume to her lips. Thrice she kissed it, thrice inclined her head, thrice she felt the delirium of acquiescence. Her ritual performed, she turned to page 1367, which gave the times of the departure of the airships from the island in the southern hemisphere, under whose soil she lived, to the island in the northern hemisphere, whereunder lived her son.

Although the action is deeply distressing, she leaves her room and travels by subway and then by airship to see him. On the airship, she is upset by direct interactions with people or the outside world; becoming quite horrified when a flight attendant touches her, and when sunlight intrudes into her cabin. When she reaches Kuno, he tells her of his experience of finding a passageway up to the outside, and of spending several days outside, although the air there is difficult to breathe. And for this excursion, he has been threatened with Homelessness, with exposure to the air, and thus with death.

Vashti is both fascinated and disgusted by his story. She returns home and resumes her ordinary routines, and in fact loses contact with her son. Not long after the distressing visit, the politics of life within the Machine evolve further. No longer can citizens access respirators to go to the surface. In addition, the worship of the Machine solidifies into a religion.

Vashti hears that he has been transferred back to somewhere near her, under Australia. A few years after her visit, he calls her and tells her, "The Machine stops." She discusses this with a friend, and wonders if he was referring to

the unharmonious noises that had been interrupting the music they listened to. There are other problems, such as jarring noises. She and her friends complain, and are told their complaints will be forwarded to the Central Committee. After a while, they cease to notice the problems of moldy fruit, of unclean bath water, of dysfunctional poetry machines, even of beds that failed to appear.

And then the communication network ceases, while Vashti is giving a lecture. She hides in her room, listening to the sounds of people fighting in the tunnel outside her room. Listening to them dying. When the sounds of battle are replaced by the sounds of death and dying, she ventures out, and finds her son. As he bleeds, he exults that they are talking, touching. And he tells her that there are people living on the surface of the earth. They spend their last living moments together, mourning the loss of the experiences they never had because they remained within the Machine. Their moment of connection is ended by an errant airship that destroys the underground city.

Precis of Ethics Theories

We offer here a very brief introduction to three major theories of ethics, and note that there are a number of good ethics textbooks (e.g., (Copp 2005; LaFollette and Persson 2013)) that offer more in-depth and insightful introductions to these and other theories. Although some ethicists engage in comparative study whose purpose is largely descriptive, nearly all work in ethics – both academically and in the wider world – is, by contrast, normative: that is, it argues how the world should be understood, and how people ought to act. Most approaches to ethics adopt one of three basic postures.

Deontological Ethics

Deontology understands ethics to be about following the moral law. In its most widely-recognized form, was developed by Immanuel Kant in the late nineteenth century, but has ancient roots in both Divine Command traditions (such as ancient Israelite religion, the source of the Ten Commandments and the basis of Judaism, Christianity and Islam) and in other legal codes. The basic question of deontology is "what is my duty?" According to deontology, that duty can be understood in the form of laws. Kant's innovation was to locate the universal law within the individual, rather than in an externally-given system; to Kant, following a set of laws imposed by another, without submitting them to the scrutiny of one's internal sense of the law, represents a moral failure of the individual. Although the theoretical rationale for law-based ethics and Kantian deontology is different, in both systems, any true law will be universally applicable.

Isaac Asimov's I, Robot books (Asimov 1950) look at the consequences of building ethics based on his Three Rules. Students may perceive deontological analysis to be analogous to application of axiomatic systems. The underlying questions become, "How are rules applied to decisions?" and "What are the right rules?" The latter question is one of mechanism design, namely, what rules do we put in place in order to achieve our desired social goals?

Virtue Ethics

Virtue ethics (also known as teleological ethics) is focused on ends or goals. The basic question of virtue ethics is “who should I be?” Grounded in classical philosophy and outlined most clearly in Aristotle, virtue ethics is organized around developing habits and dispositions that are conducive to developing the qualities or abilities that help one achieve those ends, and, by extension, to flourishing as an individual (Annas 2006). In contrast to deontological ethics, virtue ethics considers goodness in local rather than universal terms (what is the best form/version of this particular thing, in these particular circumstances?) and emphasizes not universal laws, but local norms. Virtue ethics was, for many centuries, the dominant mode of ethical reasoning in the west among text scholars and the educated classes. It was eclipsed by utilitarian ethics in the late 18th and 19th centuries, but has seen a resurgence, in the past fifty years, among philosophers, theologians, and some literary critics. For two thinkers who advance this widely-acknowledged narrative, see Anscombe (2005) and MacIntyre (2007).

Utilitarianism

The most recent approach, utilitarian ethics, was developed by Jeremy Bentham and John Stuart Mill in the late 18th to mid-19th century. The basic question of utilitarianism is “what is the greatest possible good for the greatest number?” — or, in William K. Frankena’s more recent formulation, “the greatest possible balance of good over evil (Frankena 1963).” In computer science, and broadly in the social sciences we use “utility” as a proxy for social good. The underlying assumption is that utility can be quantified as some mixture of happiness or other qualities, so that we can compare the utilities of individuals, or the utility that one person derives in each of several possible outcomes. The so-called Utilitarian calculus compares the sum of individual (dis-)utility over all people in society as a result of each ethical choice. While classic Utilitarianism does not associate probabilities on possible outcomes, and is thus different from decision-theoretic planning, the notion of calculating expected utility as a result of actions fits well into the decision-theoretic framework.

Ethical Theory in the Classroom

While all three schools have proponents among philosophers, theologians, and other scholars who work in ethics, broader cultural discourse about ethics tends to adopt a utilitarian approach, often without any awareness that there are other ways to frame ethical inquiry. This larger cultural reliance on Utilitarianism may help explain why it consistently seems, to the students, to be the most crisply-defined and “usable” of the ethical theories. But there are significant critical shortcomings to Utilitarianism, most particularly its in-substantive definition of “goodness” and the fact that it permits (and even invites) the consideration of particular problems in isolation from larger systems. These shortcomings limit our ability to have substantive ethical discussions, even insofar as everyone assents to Utilitarianism; a shared reliance on the principle of “the greatest good for the

greatest number” does not help us agree about what goodness is, or even to reach an agreement about how to define it. These same limitations surface in student conversations about ethics. One common problem in their application of Utilitarianism is that they may look too narrowly at who is affected by a given decision or action. One example is the question of whether to replace factory workers with robots. They may focus on the happiness of the factory owners, shareholders, and those who can purchase the manufactured goods more cheaply, without considering the utility of the factory workers and those whose jobs depend on factory workers having money to spend; still less are they likely to consider the shortcomings of an ethical model that makes it possible to conceive of human beings and machines as interchangeable.

In our limited experience teaching Science Fiction and Computer Ethics with primarily computer science students, we find that students are often drawn first to Utilitarianism, perhaps because it seems more computational than the others. One important aspect of the course, or any reading therein, is to broaden their experience and expose them to other modes of thinking and reasoning in this space. While we believe it is essential to help students move beyond the assumption that utilitarianism is the only approach (let alone the best approach) to ethical questions, our aim is not to demonstrate the superiority of one approach over the other, but rather to help students understand the uses and limits of each approach. Furthermore, the approaches are not necessarily mutually exclusive; Recent theorists have argued that virtue ethics is best seen as part of successful deontology (McNaughton and Rawling 2006).

Themes to Discuss

This story, written over 100 years ago, is remarkably prescient in some respects. It imagines a world of video chats and TED talks, of extremely broad, shallow social networks. We are, by comparison to the characters in Forster’s story, fortunate to still have physical lives, access to real food, and the opportunity for direct interactions. Indeed, there is a mention of “all of” Vashti’s children, but no indication of how they are conceived — except in the discussion of the passengers on the airship, including “He had been sent to Sumatra for the purpose of propagating the race.” Vashti isolates herself to sleep, and to talk to her son. After her three-minute conversation with her son, though, “Vashti’s next move was to turn off the isolation switch, and all the accumulations of the last three minutes burst upon her. The room was filled with the noise of bells, and speaking-tubes.”

While reading the story there are a number of areas for engagement with students. The goal of the course is to allow students to bring these issues to the fore themselves, but we mention a few interesting directions and tie them to the basic ethical theories above.

Always Connected

Forster anticipates the notion of continuous connection discussed by Turkle and others (Turkle 2006). Stone discusses the state of “continuous partial attention,” (Stone 2007;

2008), induced by constant connectivity (in Stone's case, via cell phone, IM, texting, etc.) Stone says that such steady stream of interruptions and distractions "contributes to a stressful lifestyle, to operating in crisis management mode, and to a compromised ability to reflect, to make decisions, and to think creatively."

When Vashti isolates herself, there are social repercussions. Sherry Turkle has written about the "tethered self", always physically attached to communication devices. Vashti is clearly tethered by the proximity of her communication apparatuses to her armchair. Although she is described as a deep thinker, we see no evidence of this except in the outline of her lecture. However, it seems impossible to cover such a range of material (the history of Australian music) in the ten minutes she speaks. Thus, we can conclude that, indeed, her ability to reflect is limited. Instead, she is called upon to issue snap judgements. If Forster had thought of the "like" button, Vashti would have been as socially obligated to respond thus to her myriad acquaintances thoughts and reviews as is the main character of David Egger's *The Circle* (Eggers 2013), unable to think new thoughts because of the demand to respond and validate others' output.

It is difficult to talk with students who sit, absorbed in their phones and other devices, often half- or less present in class, about the ethics of attention and accessibility. Casting the issue in an alien environment such as Forster's gives them a new perspective on the question.

A utilitarian analysis of constant connection can look at the balance of instantaneous rewards for connectedness and the diminished abilities of an always-connected person to work, enjoy leisure, or relate to others. This can lead to a discussion of the utility of multi-tasking, or to the social utility of loneliness and connection.

Dependent on the Machine

In the setting of the story, Vashti and the others living underground are dependent on the Machine for food, water, breathable air, entertainment, and health care. She is portrayed as being able to walk, but for no great distance. She is never shown taking care of herself or others, except by summoning a bath or bed as she desires.

One can challenge students to consider what supports their lives. We can discuss the effects of technology on mundane aspects of our lives. For instance, industrial agriculture brings us food both in re-heatable, prepackaged trays or ready to cook portions, removing us from the production of food (farming, processing, cooking). While the story supposes an extreme case of this reality, it is an interesting opportunity to discuss how much utility we derive from our own machines, and what effect the pre-packaged life style has on us as a society and as individuals.

From a utilitarian viewpoint we can argue that this dependence creates opportunities for efficiency and happiness. If everyone lives in the same manufactured blocks and has their needs met, we are maximizing utility. One area of modern life where there is active argument against this is factory farming. Many make one of two argument here, one deontological and one utilitarian in nature. The deontological argument is that animals have rights and we should respect

those. The utilitarian is that we should consider animal welfare when we do our maximization.

A higher level question can be leveled at the purpose of the machine. As technologists we are building things and we should consider what the apparatus of *our* machine looks like; where we place value while doing research and development. Using apps to constantly ratchet up our hedonic set points, so we can never be without entertainment or communication, one can argue, as Pressler does, that we are dependent on a machine that is not designed to do anything more than making us fleetingly happy (Pressler 2014). Other questions of dependence on the machine can raise the question *for whom* do we make technology. There are other good exemplars of this topic which should also be considered in a good course, such as Charlie Chaplin's *Modern Times* — not strictly science fiction, but an excellent early warning about the problems of mechanizing the workforce. This allows the students to discuss ethical questions beyond utilitarianism: What do we *want* the machine to do? What *rules* should we follow when building the machine itself?

Physically Isolated

It is clear, as we read the story, that Vashti is more physically isolated than is strictly necessary for her society. She turns down an invitation to visit a public garden. When she does step outside her room, there is public transportation available, and at least one other person on the subway with her. There are many others on the airship, although each gets their own cabin. (In that way, airship transport more closely resembles ocean liners than modern airplanes.) Vashti's life is lived primarily from her armchair, from whence she can speak via speaking tube and imperfect image to individual friends or to a lecture audience of people in their own, identical armchairs.

Vashti is described as "a swaddled lump of flesh" in the first paragraph. Later, she is described as having no teeth. She is clearly not physically fit. There is a healthcare apparatus that can appear as needed or demanded, to care for her. Modern American culture emphasizes exercise as a virtue. A Utilitarian view might conclude that exercising is an ethical choice because it diminishes dependence on an expensive and overtaxed healthcare system; increases personal utility through endorphins; increases attractiveness, and thus indirectly increases personal utility. It is not clear whether any of these apply in Forster's imagined world.

We are told that Vashti has multiple children, but there is no discussion of relationships or attractiveness. Perhaps she was inseminated. We do know that she gave birth. "But she thought of Kuno as a baby, his birth, his removal to the public nurseries, her own visit to him there, his visits to her — visits which stopped when the Machine had assigned him a room on the other side of the earth. 'Parents, duties of,' said the book of the Machine, 'cease at the moment of birth. P.422327483.' True, but there was something special about Kuno — indeed there had been something special about all her children".

Vashti may struggle with depression. When a lecture she gives is poorly received, she wishes to die and requests euthanasia. However, deaths must be balanced with births and

she has always been denied. Certainly, she does not have exercise as a palliative to her bleak world. However, Kuno speaks of “dietary tabloids” (food pills). It is possible that citizens are medicated for depression, though this is not stated explicitly.

There seem to be no rules about personal care in this world. Vashti is rule abiding. “And if Kuno himself, flesh of her flesh, stood close beside her at last, what profit was there in that? She was too well-bred to shake him by the hand.” It seems that she lives a deontologically sound life, according to her society. One can discuss, however, whether those rules and social strictures are appropriate for our society.

There is an interesting question to discuss about whether Vashti cares about Kuno, and whether she cares for (takes care of) him. She taught him to function, and she eventually goes to him when he asks her to. But once there, she can barely hold a conversation with him, and cannot touch him — until he is lying bloody and dying on the ground, on the last page of the story. What does it mean for her to care for him? For her other acquaintances?

Ethical Inquiry

In addition to a surface analysis of the main themes, the real strength of “The Machine Stops” (and many other works of science fiction besides) is that it rewards multiple lines of ethical inquiry. We outline below how each of the three theories of ethics described above can provide a way into exploring the ethical questions raised by the story.

As noted above, utilitarianism can be a helpful way to think about the costs and benefits of a technologically-dependent life, both in Forster’s story and in our own lives. We read Forster’s story as a fierce critique of utilitarian thinking. The Machine seems to address all our needs, when formulated in a utilitarian way: it provides the most good for most people, most efficiently, by defining goodness as itself efficiency. But this efficiency, in which all experiences are streamlined, has led to a sterile uniformity, and removed the possibility of meaningful choice. In the world of the Machine, people are moved halfway across the world because a room has become available for them there, or because they are required as breeding stock in that particular locale, and their consent does not seem to matter; on the other hand, they lose very little by moving, since all cities are the same and most social contact takes place through a disembodied network. It is a world with no room for the sort of experience that Kuno finds meaningful. It does not furnish him with the sensory experiences he wants, and persecutes him for his efforts to seek such experiences outside its boundaries. One could give a utilitarian reading of the story by arguing that the notion of “best” that informs the machine is the problem; it is incorrectly or improperly defined, rather than the fault of utilitarianism itself.

A deontological approach to the story invites us to consider the different ways of understanding laws, and the conflict between socially-instituted laws and “natural” or internally-grounded laws, and to use the principle of universality to adjudicate the conflict. Vashti accepts the dicta of the prevailing culture — even seeming to worship the Machine itself and its Book, with its explanations of duties —

whereas Kuno rejects them, as if in response to an internally-known moral law such that Kant describes. Forster makes clear that the Machine’s law, and its reach, seems to be universal; he invites the reader to reject the Machine, and the values that follow from living within it, by showing us its limits from the perspective of the visionary Kuno, who tells his mother that “The Machine is much, but it is not everything.” The Machine purports to cover the whole survivable world, but as Kuno discovers, the surface of the earth offers much more; as the end of the story demonstrates, the Machine cannot be relied upon to fulfill even its own promises of endless undisturbed function. The Machine, however powerful, has limits, and is contrasted with the more truly universal power of sensual experience. Vashti comes to know this more universal law as well as the story ends, as she joins her on in grieving for “the sin against the body - it was for that they wept in chief; the centuries of wrong against the muscles and the nerves, and those five portals by which we can alone apprehend.” The experience of touch, which Kuno seeks in his attempted escape, and then experiences again as he and his mother meet for the last time in the bowels of the dying Machine, is portrayed as more fundamental and universal than anything the Machine can offer, even though the systems of the machine are designed to suppress human contact.

A virtue ethics reading of the story puts a different slant on the clash between Vashti and her son. Although the basic events of the story can be accounted for by Kuno’s rejection of the laws of the Machine, in contrast to his mother’s submission to them, we can draw a richer sense of Forster’s ethical arguments — and their direct connection to our own lives — by thinking about the different visions of personhood that Kuno and Vashti use to make their choices. Vashti, who subscribes to the ideals of the Machine, is guided in her thinking not only by questions of what is legal, but of what is not the proper thing, it is not mechanical, it is not “decent-mechanical” — this term is never defined, exactly, but its scope is made clear when Vashti affirms in the same breath that “it is perfectly legal, perfectly mechanical” to visit the earth according to proscribed procedures, and when Kuno admits that “it is not the proper thing, it is not mechanical, it is not decent” to go exploring as he has done, even though such explorations are technically legal. To be mechanical is bigger than following the laws; it is about adhering to the norms of the Machine. Vashti, who has lived by these norms for many years, has absorbed even the sensory horror of light and touch that the Machine has trained her to have. By contrast, Kuno is able to challenge those norms — even though they have shaped his entire life — because he conceives of a different idea of what human flourishing looks like. His idea that humans were made to be different than the Machine has made them leads him to adopt new ways of living, which in turn helps him refine his new definition of human flourishing. This process is clear in Kuno’s explanation to his mother, when he tells her that “we have lost a part of ourselves. I determined to recover it, and I began by walking up and down the platform of the railway station outside my room. Up and down, until I was tired, and so did recapture the meaning of near and far.” Kuno walks until his

strength increases, which in turn allows him to explore further within the Machine, leading him to imagine the possibility of getting outside of it. In exercising the parts of himself that “mechanical” thinking deems unimportant, Kuno not only “develop[s] my sense of space and my muscles,” but also begins to develop new goals, which in turn push him to cultivate his body, his thoughts and his senses. Virtue ethics thus enables us frame the clash between Vashti and Kuno as a clash between different notions of what a person should be, different teleological visions that call for different daily practices and different ways to evaluate even the smallest experiences.

Conclusions

We have used this story in our Science Fiction and Computer Ethics course (Burton, Goldsmith, and Mattei 2015). Because we teach the course through discussion rather than lecture, the actual topics covered, the ethical theories applied, and the analyses vary from year to year. However, we have found it a potent introduction to issues of over-connection and isolation, and a case study for ethical analysis. In addition, this story offers support to the argument that science fiction has much to teach us about possible future technology — both ideas for development and warnings of possible consequences of the use of the technology.

Acknowledgements

Data61 (formerly known as NICTA) is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

References

- Annas, J. 2006. Virtue ethics. In Copp, D., ed., *The Oxford Handbook of Ethical Theory*. Oxford University Press.
- Anscombe, G. E. M. 2005. Modern moral philosophy. In Geach, M., and Gormally, L., eds., *Human Life, Action and Ethics: Essays by G. E. M. Anscombe*, St. Andrews Studies in Philosophy and Public Affairs. Imprint Academic.
- Asimov, I. 1950. *I, Robot*. HenryBennet.
- Bates, R. A. 2011. AC 2011-1669: AI & SciFi: Teaching writing, history, technology, literature and ethics. In *ASEE Annual Conference & Exposition*.
- Bauer, J. 2003. The challenge to international human rights. In Monshipouri, M.; Englehart, N.; Nathan, A. J.; and Armonk, K. P., eds., *Constructing Human Rights in an Age of Globalization*. M.E. Sharpe.
- Bean, J. C. 2011. *Engaging ideas: The professor’s guide to integrating writing, critical thinking, and active learning in the classroom*. John Wiley & Sons.
- Bonnefon, J.-F.; Shariff, A.; and Rahwan, I. 2015. Autonomous vehicles need experimental ethics: Are we ready for utilitarian cars? *arXiv preprint arXiv:1510.03346*.
- Boyer, E. L. 1996. The scholarship of engagement. *Journal of Public Outreach* 1(1):11–20.
- Boyer, E. L. 1997. *Scholarship Reconsidered: Priorities of the Professoriate*. Jossey-Bass.
- Burton, E.; Goldsmith, J.; and Mattei, N. 2015. Teaching AI ethics using science fiction. In *1st International Workshop on AI, Ethics and Society, Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Burton, E. 2014. *Fantasy and Responsibility: Phronesis and ethical meaning-making in the Chronicles of Narnia*. Ph.D. Dissertation, University of Chicago. <http://search.proquest.com.proxy.uchicago.edu/docview/1615100120>.
- Colbeck, C. L. 1998. Merging in a seamless blend: How faculty integrate teaching and research. *The Journal of Higher Education* 69(6):647–671.
- Copp, D. 2005. *The Oxford handbook of ethical theory*. Oxford University Press.
- Cummings, M. L., and Ryan, J. C. 2014. Shared authority concerns in automated driving applications.
- Davis, B. G. 2009. *Tools for Teaching*. Jossey-Bass, second edition.
- Dils, L. S. 1987. Science fiction and the future, Yale-New Haven Teachers Institute. <http://www.yale.edu/ynhti/curriculum/units/1987/2/87.02.04.x.html>.
- D’Onfro, J. 2015. The huge, unexpected ethical question that self-driving cars will have to tackle. *Business Insider*. <http://www.businessinsider.com/the-ethical-questions-facing-self-driving-cars-2015-10>.
- Eggers, D. 2013. *The Circle*. Knopf Canada.
- Frankena, W. K. 1963. *Ethics*. Prentice-Hall.
- Garcia Iommi, L. 2011. Let’s Watch a Movie!: Using film and film theory to teach theories of international politics from a critical perspective. *American Political Science Association 2011 Annual Meeting*.
- Goering, S. 2014. Using children’s literature as a spark for philosophical discussion: Stories that deal with death. In Mills, C., ed., *Ethics and Children’s Literature*. Ashgate.
- Goldsmith, J., and Mattei, N. 2011. Science fiction as an introduction to AI research. In *Proc. 2nd AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI)*.
- Goldsmith, J., and Mattei, N. 2014. Fiction as an introduction to computer science research. *ACM Transactions on Computer Science Education* 14(1):4:1–4:14.
- Goodall, N. 2014a. Ethical decision making during automated vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board* (2424):58–65.
- Goodall, N. J. 2014b. Machine ethics and automated vehicles. In *Road Vehicle Automation*. Springer. 93–102.
- Haworth, J. G., and Conrad, C. F. 1995. Curricular transformations: Traditional and emerging voices in the academy. In Haworth, J. G., and Conrad, C. F., eds., *Revisioning Curriculum in Higher Education*. Simon & Schuster Custom Publishing. 191 – 202.
- Hoffman, M. E.; Dansdill, T.; and Herscovici, D. S. 2006. Bridging writing to learn and writing in the discipline in computer science education. In *Proceedings of the 37th*

- ACM Technical Symposium on Computer Science Education (SIGCSE 2006), 117–121.
- LaFollette, H., and Persson, I. 2013. *Blackwell Guide to Ethical Theory*. Wiley-Blackwell.
- MacIntyre, A. 2007. *After Virtue: A Study in Moral Theory*. University of Notre Dame Press.
- McLeod, S. H., and Soven, M. 1992. *Writing Across the Curriculum*. Sage Publications.
- McNaughton, D., and Rawling, P. 2006. Deontology. In Copp, D., ed., *The Oxford Handbook of Ethical Theory*. Oxford University Press.
- Merry, S. E. 2003. Human rights law and the demonization of culture (and anthropology along the way). *PoLAR: Political and Legal Anthropology Review* 26:1555–2934.
- Mihail, P.; Rubin, B.; and Goldsmith, J. 2014. Online discussions: Improving education in CS? In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education (SIGCSE 2014)*.
- Nussbaum, M. 1990. *Loves Knowledge: Essays on Philosophy and Literature*. Oxford University Press.
- Pease, A. 2009. Teaching ethics with science fiction: A case study syllabus. *Teaching Ethics: The Journal of the Society for Ethics Across the Curriculum* 9:75–82.
- Perry, W. G. 1980. Cognitive and ethical growth: The making of meaning. In Chickering, A. W., and Associates., eds., *The Modern American College*. Josey Bass. 76–109.
- Pressler, J. 2014. Let’s, like, demolish laundry. *New York Magazine*.
- Rodwell, G. 2013. *Whose History?: Engaging History Students Through Historical Fiction*. University of Adelaide Press.
- Rogaway, P. 2015. The moral character of cryptographic work. Cryptology ePrint Archive, Report 2015/1162. <http://eprint.iacr.org/>.
- Stone, L. 2007. Living with continuous partial attention. *Harvard Business Review* 85(2):28.
- Stone, L. 2008. Continuous partial attention—not the same as multi-tasking. *Business Week*.
- The Executive Board of the American Anthropological Association. 1947. Statement on human rights. *American Anthropologist* 49(4):539–543.
- Turkle, S. 2006. Always-on/always-on-you: The tethered self. In Katz, J., ed., *Handbook of Mobile Communication and Social Change*. MIT Press.
- Walker, H. M. 1998. Writing within the computer science curriculum. *ACM SIGCSE Bulletin* 30(2):24–25.
- Wallace, A. 2015. Who won science fiction’s Hugo awards, and why it matters. *Wired Magazine*, [Online; accessed 22-October-2015]. <http://www.wired.com/2015/08/won-science-fictions-hugo-awards-matters/>.
2015. Why self-driving cars must be programmed to kill. *MIT Technology Review*. www.technologyreview.com/view/542626/why-self-driving-cars-must-be-programmed-to-kill/.