What do Trollies Teach Us About Responsible Innovation?

Forthcoming chapter in Charles Tandy (ed.), *Death And Anti-Death, Volume 19: One Year After Judith Jarvis Thomson (1929-2020)*, Ann Arbor, MI: Ria University Press. *Forthcoming*. ISBN 978-1-934297-35-3

CHAPTER XYZ

What do Trollies Teach Us About Responsible Innovation?

Steven Umbrello

In her 1967 paper "The Problem of Abortion and the Doctrine of the Double Effect", Philippa Foot first laid out clearly the philosophical dilemmas that emerge when we consider the concepts of *intention* and *foresight* in tandem. In this paper, Foot presents an example of a surgeon who, while performing surgery on a mother in labor, must choose between the following two means of attempting to save her life: (1) perform a hysterectomy, *intending* to remove the uterus while *foreseeing* that an additional consequence will be the death of the fetus or (2) perform a late-stage abortion, directly intending to end the life of the fetus. The literature now contains many variations of cases with this basic structure. Almost a decade later, Judith Jarvis Thomson would come to call these cases *trolley problems*.[1]

Following these early discussions of trolley problems, an entire literature has emerged that discusses the relationship between moral intuitions, intentions, and our ability, if any, to foresee the consequences of our actions. Much work in moral philosophy has the aim of providing clarity as to what we ought to do in these cases. Beyond the walls of the university, popular culture has also made great use of such dilemmas. Consider, for example, Batman's tragic choice

[1] Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), 204-217.

between saving the public hero Harvey Dent or his love interest Rachel Dawes in *The Dark Knight* (2008), or Captain Kirk's attempt to successfully complete the no-win *Kobayashi Maru* training exercise.[2] In many cases, these moral choice architectures (i.e., designed environments where moral choices are necessary) are used by writers to tease out their audience's moral intuitions. Among fans, there is no consensus as to whether Bruce Wayne should have saved Harvey Dent; foresight is often not 20/20.

These trolley problems, as we will explore further in this chapter, are not only found in the academic literature and works of fiction. In the age of ever greater automation, artificial intelligence and big data, we are continually being confronted with these scenarios in the real world. Autonomous vehicles (AVs) have arrived and are here to stay. AVs have already been successful in reducing congestion, and even demonstrating greater navigational efficacy then human drivers.[3] The early use of AVs has also revealed some of their weaknesses, betraying some of their fundamental faults and our tenuous ability to maintain meaningful control over them.[4] Long gone are the days of

---

[2] Meyer, Nicholas (Director). (1982). *Star Trek II: The Wrath of Khan* [Film]. Paramount Pictures.

[3] Chen, B., Sun, D., Zhou, J., Wong, W., & Ding, Z. (2020). A future intelligent traffic system with mixed autonomous vehicles and human-driven vehicles. *Information Sciences*, *529*, 59-72.

[4] Calvert, S. C., Heikoop, D. D., Mecacci, G., & van Arem, B. (2020). A human centric framework for the analysis of automated driving systems based on meaningful human control. *Theoretical Issues in Ergonomics Science*, *21*(4), 478-506. See also Umbrello, S., & Yampolskiy, R. V. (2021). Designing AI for explainability and verifiability: a value sensitive design approach to avoid artificial stupidity in autonomous vehicles. *International Journal of Social Robotics*, 1-10.

merely worrying about a bifurcating trolley track; we must now decide how AVs *can* and *should* make decisions in such choice architectures Our sociotechnical world in which technological artifacts have become inextricably linked with our lives, organizations, institutions, and policies has created a moral quagmire. Who, if anyone, is morally responsible if an AV crashes into a wall in order to avoid hitting a jaywalking pedestrian, thereby resulting in the death of its driver?

These are difficult questions to ask and even more difficult ones to answer. This demonstrates that thinking in terms of trolley problems is widespread in our sociotechnical world. There are further, perhaps more fundamental, questions with the same kind of structure. In light of our diminished control over our technological creations, getting clearer on who is responsible for their design, and *how* they are designed, is pivotal. Shedding light on these questions is the guiding aim of this chapter. Rather than contributing to the extensive literature focusing directly on the trolley problem, this chapter explores the value of trolley cases in what has become known as *Responsible Innovation* (RI). At its core, RI aims to prevent tragic choice scenarios from arising, using techniques like salient, value-oriented design. To this end, this chapter provides some crucial background on RI, discusses the importance of centralizing human values in design, as well as demonstrating the value of trolley thinking in RI more broadly.

Responsible Innovation

It is hard to deny the benefits of technological progress. Examples include new medicines, clean drinking water, global and instantaneous communications, as well as access to information spanning all recorded history. In turn, these innovations have ushered in new economic systems that have led to economic prosperity. Non-economic indicators, like life expectancy and overall health, also reveal continual progress. High infant mortality rates, once normalized, have decreased globally. Despite these benefits, we would be hesitant about claiming that technological innovations are goods in and of themselves. Rather, when presented with a particular technological innovation, we must ask: "is *this* technology good?" The need to ask this question results from the realization that there have been many once-promising innovations that ultimately raised challenging moral questions. Asbestos, a fibrous silicate mineral that was touted as an attractive electrical insulator, was later revealed to be a toxic, cancer-causing substance when inhaled. Asbestos is not alone in moving from widespread use to obsolescence. After tragic events like the 1937 Hindenburg disaster, transportation craft like zeppelins ceased being appealing, other than in the popular imagination.

It is for reasons such as these that global institutions are actively seeking to orient innovation towards addressing these moral issues head on and at a global scale. The seventeen United Nations Sustainable Development Goals (SDGs) are examples of this kind of orientation. The SDGs have the explicit aim of developing resilient infrastructure, as well as promoting sustainable industrialization and innovation (SDG #9).[5] The SDGs are not separable, nor are

[5] United Nations. (2019). Sustainable development goals. In *GAIA* (Vol. 28, Issue 2, p. 73). https://doi.org/10.14512/gaia.28.2.1

they rank-ordered in terms of their importance. Instead, they are developed to be co-constitutive and for progress towards one goal to co-vary with progress towards the others. Here is just one example. SDG #9 (building resilient infrastructure and promoting sustainable industrialization) mutually reinforces the achievement of other goals, such as ensuring access to affordable and clean energy (SDG # 7) and the taking of climate action (SDG #13). Innovation understood in this manner is no longer understood as developing bigger and better vehicles but instead as innovating towards better futures that we can pass on to future generations. An orientation at the global scale towards the solving of existent technological challenges, coupled with an orientation that anticipates and designs *for* human values rather to their detriment, is crucial if innovation is to be carried out responsibly.

Although the discussion of values is situated in a long philosophical tradition, it remains inseparable from discussions of technology. Novel technologies include modular prosthetic limbs[6] that can return functionality to the previously impaired and neuro-enhancing technologies, like the promised Neuralink[7], which may one day lead to a brain-machine interface. Although the latter may be far off,

[6] Hotson, G., McMullen, D. P., Fifer, M. S., Johannes, M. S., Katyal, K. D., Para, M. P., ... & Crone, N. E. (2016). Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. *Journal of neural engineering*, *13*(2), 026017.

[7] Pisarchik, A. N., Maksimenko, V. A., & Hramov, A. E. (2019). From novel technology to novel applications: Comment on "An integrated brain-machine interface platform with thousands of channels" by Elon Musk and Neuralink. *Journal of medical Internet research*, *21*(10), e16356.

whether technologies like these are morally acceptable will depend on the specifics of their designs given that those design choices will have impacts across space and time. At the core of this issue is the concept of *value-ladenness*. According to this line of thought, technologies are never value-neutral and are better understood as embodying the values of their creators. If we are responsible for our innovations, in virtue of these technologies inheriting and instantiating our values, then responsible design is more important than ever. It will be worthwhile to first take a closer look at some low-tech examples of value-ladenness.

Commissioned by Constantine the Great after his mother visited Bethlehem, the Basilica of the Nativity was completed almost two hundred years after its ground-breaking in c. 565 C.E.[8] One of the stunning features of this building is the main entrance, a doorway measuring less than five feet in height. This doorway, aptly called the "the Door of Humility", forces visitors to enter the basilica bowed. However, the door was not always this low, meaning that entering has not always required bowing. The original crusader doorway can in fact still be seen above the more recent, shrunken, doorway that was built during the Ottoman period. Rather than enforcing humility, the aim of lowering the door was in fact to prevent thieves from entering the basilica on horseback, or with carts, that they could then use to quickly abscond with the church's treasure.[9]

---

[8] Madden, A. M. (2012). "A Revised Date for the Mosaic Pavements of the Church of the Nativity, Bethlehem". *Ancient West and East*. **11**: 147–190.

[9] Rees, M. (2002, May). *The Saga of the Siege*. Time. http://content.time.com/time/subscriber/article/0,33009,1002452,00.html.

The basilica's security feature is not the only example of low-tech innovation that we can draw on. One of the most famous examples often discussed in the literature on value-ladenness was first introduced by Langdon Winner in his 1980 paper "Do artefacts have politics?".[10] Winner took as a case study the low-hanging overpasses that were built across Long Island, NY at the beginning of the twentieth century by the architect Robert Moses. He argued that Moses designed the low overpasses with the intention of preventing public transit buses coming from the inner cities from accessing his favourite beaches. As a result, those who depended on public transportation, primarily the largely African American urban poor, were unable to reach those shores. Access to the beaches was available only to the upper- and middle-class white citizens who could afford cars. The bridges, which stand to this day, were designed to embody the values (or in Winner's words "politics") of their creator Robert Moses; in this case, these were racist values.

Value-ladenness has been at the forefront of the ethics of technology since the 1980s and forms the bedrock of what has since come to be called the *design turn in applied ethics*.[11] This design turn has stressed both the importance of the concept of value-ladenness and the responsibility of designers in their innovating. Moving beyond low-tech examples, it is possible to see how novel information and communication technologies embody values to an even greater degree than their low-tech counterparts. Financial

[10] Winner, L. (1980). Do artefacts have politics? *Daedalus*. *109*(1), 121-136.

[11] Van den Hoven, J., Miller, S., & Pogge, T. (2017). The design turn in applied ethics. *Designing in ethics*, 11-31.

and insurance algorithms, medical diagnostic systems and even geographical information systems are built using models and algorithmic processes that are difficult, if not impossible, for the vast majority of their users to understand. Crucially, these technologies influence and modify how we interact with each other and with the technologies themselves. Without systematic and exacting assessments of what values are being embodied and how, the pervasiveness of novel technologies creates the conditions for the likes of Robert Moses to pursue their less-than-noble aims.

Remaining cognizant of this risk, the twin aims of the designer are to ensure that new innovations embody our shared moral values and to use these innovations to address the obstacles that we face as a global community. Achieving these aims requires an anticipatory approach to design. The technological systems that are currently becoming ubiquitous are transformative technologies such as AI, nanotechnologies, and biotechnologies. The potential impact of such technologies is too great to permit a reactive, rather than proactive or anticipatory, approach. This, of course, does not come without its difficulties. We find ourselves, even when we take a value sensitive design approach to novel technologies,[12] with a diverse array of moral values that we want systems to embody. In many cases, we are faced with a greater number of moral commitments than the design scenario permits us to effectively balance. To use the terminology of Jeroen van den Hoven et al., (2012), these situations involve *moral overload*. Moral overload can occur not only when we are overwhelmed with many relevant

[12] Friedman, B., & Hendry, D. G. (2019). Value sensitive design: Shaping technology with moral imagination. MIT Press.

moral commitment, but also in situations where two moral commitments are in genuine tension with one another.

To describe a situation as involving moral overload may be taken to suggest that the problem is intractable. However, apparent conflicts in values are sometimes exactly that: *apparent*. Moral overload should not be seen as involving an understanding of values as standing in direct opposition to one another, but more weakly as being in tension with one another. This tension can often lead to creative solutions in design and innovation, amplifying (rather than merely *balancing*) the values in question. This will be explored in more detail when discussing the engineering modality of trolley cases.

One example is *Fairphone*, an Amsterdam-based company founded by Dutch entrepreneur Bas van Abel. *Fairphone* aims to develop sustainable smartphones, ensuring that the environmental impact of the phone's production is minimized through responsible sourcing (Homerun.co., 2018). Gold and copper are used in smartphone PCB circuit boards and often have supply chains that fare poorly when it comes to sustainability. *Fairphone* sources their gold from *Gold Circuits Electronics Ltd.*[13] who responsibly source their gold and exclusively use recycled copper. Moral values like sustainability and fairness, along with economic values like profitability, are balanced as part of an integrated design.

The notoriously fog-filled city of Milan offers another example of multiple values being integrated as part of the

[13] Fairphone. (2021, May 6). *Our Impact*. Fairphone. https://www.fairphone.com/en/impact/?ref=header.

design process. The now famous *Bosco Verticale* (vertical forest) is a pair of twin condominium towers in the heart of the city. The towers contain over 900 trees, 5,000 shrubs, and 11,000 perennial plants as well as 8,900 square metres of terraces.[14] In addition to providing high-rise housing in a densely populated city, the towers produce oxygen while simultaneously mitigating the effects of the city's smog. Values like sustainability, beauty, efficiency as well as profit are all accounted for in a single design.

These examples demonstrate not only the balancing of multiple values in design, but also that innovation itself can be understood as a moral concept as well as a technical one. More precisely, a moral understanding of innovation requires us to embody and promote our moral obligations and thereby forces us to seek novel and creative ways of satisfying those obligations.[15] This does not, of course, mean that we will *actually* be able to embody every relevant moral value in every instance of design. It does, however, motivate us at a conceptual level to discover means of doing so. This orientation towards finding solutions to the various moral problems is defining of responsible innovation.

This, of course, is not the only aspect of RI that is relevant here. RI can also be described as processual, meaning that we can engage in activities that actively

[14] PERI GmbH. (2013, November 14). *Il Bosco Verticale*. Il Bosco Verticale, Milan, Italy - Projects - PERI. https://web.archive.org/web/20131207052855/http://www.peri.com/en/projects/projects/skyscrapers-towers/bosco_verticale.cfm.

[15] Van den Hoven, J., Lokhorst, G. J., & Van de Poel, I. (2012). Engineering and the problem of moral overload. Science and engineering ethics, 18(1), 143-155.

enhance or diminish our *ability* to be response-*able*. While engaging at length with the vast literature on responsibility would take us too far afield, it is worth noting some of the criteria we use to assess responsibility: epistemic access and understanding, motive, volition, causality, and ability. Think, for example, of some claims often made by people attempting to minimize their responsibility: "I didn't know" (knowledge), "I didn't intend to do that" (intention), "they made me do it" (coercion), "it wasn't my fault" (causality), "I couldn't have done otherwise" (ability). Within the context of RI, we are active agents that are capable of enhancing our abilities to be responsible for our actions as well as being capable of minimizing the ability of others to *hold us* responsible. It is an explicit aim of RI to expand the set of moral problems that can be solved via design and, as a result, enhance and amplify our ability to be maximally responsible for those innovations.

But where should we situate the so-called trolley problem in relation to RI? It is not *prima facie* obvious that trolley-like scenarios can help us understand the role of responsibility in the broader discussion of RI. As previously discussed, the philosophical literature on trolley problems has ballooned into its own burgeoning sub-field. As a consequence, many common variations on the problem are well-worn and may be familiar. For the purposes of the present discussion, it is nevertheless worthwhile to discuss the basic permutations of the scenario. This will help us to better understand how the trolly problem and its relatives can help us to think about responsibility from the engineering perspective.

The standard form of the trolley scenario goes as follows (See Figure 1). Imagine that a person finds themselves at a

forking railway track, complete with a railroad switch. Barreling down the track is a train. If the person does nothing (i.e., does not flip the railroad switch), the train will continue down its current path and kill four people who are tied to the track. However, if the person does decide to pull the switch, the train will be diverted and kill only one person who is tied to the second track.

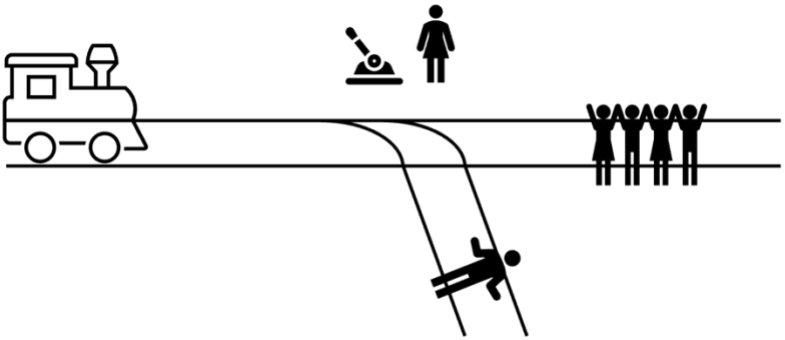

Figure 1. Trolley Scenario 1

This scenario is often presented in Philosophy 101 classes in order to tease out the moral intuitions of students, resulting in tense debate as to what one ought to do. There is some empirical evidence suggesting that the majority of people judge that it is permissible to flip the switch. In a series of experiments carried out by psychologists at Michigan State University, 90% of participants concluded that it was permissible to kill the one to save the four.[16]

[16] Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional "trolley problem". *Emotion, 12*(2), 364–370. https://doi.org/10.1037/a0025561

When polled, a majority of philosophers (69.9%) shared this view.[17]

As mentioned above, the trolley problem allows for limitless variation, including very specific scenarios (for example, scenarios where the person tied to the second track is a loved one). The judgements that people make have been shown to vary depending on arithmetically irrelevant features of the situation. One common variation involves the subject standing on a bridge under which the barreling train will pass (see Figure 2). The subject must decide whether to push a particularly large man from the bridge. Doing so will prevent the train from striking the people tied to the track but will result in the death of the man pushed from the bridge.

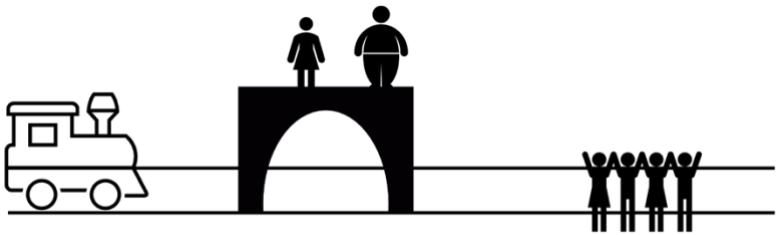

Figure 2. Trolley Scenario 2

In this slightly modified case, the consequences of (in)action remain the same. On the assumption that subjects make judgements in the initial scenario by following a consequentialist line of thinking, it seems to follow that they ought to make the same judgement in arithmetically identical variations. Surprisingly, however, this does not seem to be the case. Brain imaging studies suggest that the impersonal

[17] Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe?. *Philosophical studies*, *170*(3), 465-500.

nature of the initial scenario is responsible for respondents judging that it is permissible to flip the switch, whereas the more visceral interaction required to push the man off the bridge often results in a different judgement being made despite the calculations remaining unchanged.[18]

Regardless of the purported psychological explanations of the divergent judgements, it still may not be clear what these types of cases add to discussions of RI. Here, then, is the start of an answer to this question. Trolley scenarios are useful thought experiments that help us to conceptualize the often-nuanced landscape of moral responsibility, particularly in a complex sociotechnical world. Autonomous vehicles offer a particularly instructive example of how trolley scenarios map on to dilemmas arising from novel technologies.

So-called 'true' AVs, those that require no driver monitoring the road and allow for the driver to fully remove their hands from the steering wheel (if indeed the vehicle has one), remain beyond our reach. Nevertheless, there are extant AVs with a level of autonomy sufficient to warrant discussions of responsibility for any harmful consequences resulting from their use. It appears inevitable that these vehicles, like human drivers, will encounter choice scenarios not unlike the standard trolley cases. Imagine, for example, that you are sitting behind the wheel of an AV moving towards a pedestrian using a crosswalk. The AV's system

[18] Adams, Jessica; Frankenstein, Andrea; Alabisa, James; Robinson, Tyler; Alloway, Tracy; and Lange, Lori, "Investigating the Effects of Stress on Cognitive and Emotional Moral Decision Making" (2016). Human Factors and Applied Psychology Student Conference. 3. https://commons.erau.edu/hfap/hfap-2015/posters/3

realizes that, either in virtue of faulty braking mechanics or rapid environmental changes, the vehicle is unable to stop in time. The system needs to choose between (1) colliding with the pedestrian and (2) sharply changing its trajectory and moving into the oncoming lane, colliding with a car carrying passengers. The former results in the death only of the pedestrian while the latter may well result in the death both of the AV's driver and the passengers of the oncoming car. The choice architecture of the AV will depend entirely on the programming decisions made by its designers. An obvious issue concerns the fact that the designers make these moral decisions in their abstract, generalized, form rather than in the highly dynamic scenarios in which the decisions are actualized by the AV. If the designers take a consequentialist approach then, in situations like the one described above, the AV is likely to maintain its current course and strike only the pedestrian. If the designers choose to incorporate their own economic incentives, it may be of relevance that few consumers would purchase an AV that would sacrifice its passenger if the consequentialist calculation demanded that it do so. By contrast, drawing on the resources of virtue ethics, deontology, or some other non-consequentialist approach may rule out such a calculus altogether.

What these choice scenarios mean for RI is that designers must anticipate the potential consequences of these emergent choice structures and design systems in such a way as to maximize their ability to *take responsibility for the responsibility of others.*[19] This means taking responsibility

[19] Srivatsa, N., Kaliarnta, S., & Groot Kormelink, J. (Eds.) (2017). Responsible innovation: From MOOC to book. Delft University of Technology.

not only for those individuals and groups affected in the near future but also for future generations. The example of nuclear power is instructive. The urgency of addressing the climate crisis, coupled with our ever-increasing consumption needs, has led to nuclear power being proposed as a viable means of balancing these pressing considerations. Nevertheless, despite its potential to reduce pollution in the short term while meeting our consumption requirements, nuclear power condemns future generations to becoming shepherds of toxic nuclear waste with no clear way of managing this waste over the long term. A framing of nuclear power in this way does not seem to meet the distal responsibility requirements underpinning RI. Given that what seems responsible proximally may not be responsible distally, it is vital to not lose sight of the temporal dimension of RI.

Returning to trolley scenarios, this time from the perspective of designers and engineers, promises to shed light on the response-*abilities* of such agents in complex and dynamic sociotechnical worlds. As a jumping off point, it's striking that when presented with trolley scenarios, engineers tend to respond differently to philosophers. In particular, engineers tend to criticize the thought experiment as unrealistic.[20] The primary complaint is that the scenario as described, with a runaway train and multiple people tied to multiple train tracks, is (to put it mildly) vanishingly rare. A failure of this type is identified by engineers as a mechanical failure that should have been avoided by putting into place measures such as early warning alarms.[21] In essence, this

[20] CVL Engineers Inc. (2021, January 8). *An Engineer's Perspective on the Trolley Problem*. CVL Engineers Inc. https://cvl-eng.ca/trolley-problem/.
[21] Ibid.

response goes, the trolley problem results from poor engineering. The person who finds themselves at the switch should not have been confronted with that choice in the first place.

A philosopher may bristle at this response, given that the scenario was not constructed with this type of technical critique in mind. Despite this, the idealized nature of the thought experiment is what designers tend to rally against. Designers and engineers tend to think of ways to improve previous designs, something that is not possible within the constraints of the thought experiment. In this particular case, we are faced with the intricacies of railway safety and transportation. Whereas a philosopher may accept the set-up of a thought experiment and reason within its constraints, engineers tend to focus on what led up to the choice being described, in order to determine how design can be used to mitigate such dilemmas from arising in the future. These design histories are important in the world of RI, given that technologies do not emerge *ex nihilo* but are rather the products of innumerable human decisions that have resulted in the relatively stable designs that we see today.[22]

Our smartphones offer a particularly clear example of a design history. A new iPhone is released roughly every twelve months. Although new models are rightfully criticized as being insufficiently revolutionary when compared to their immediate predecessors, the difference between the 1st generation iPhone (released in 2007) and the latest iPhone 12 Pro (released 2020) is night and day. What this means is that we should not discount the evolutionary

[22] Van den Hoven, J., Miller, S., & Pogge, T. (2017). The design turn in applied ethics. *Designing in ethics*, 11-31.

augmentations of what designers decide to change/include in their designs in new iterations. These design decisions support and constrain what options are left open for future designers when it is their turn to sit at the drawing board. As a result, the notion of taking responsibility for the responsibility of others is crucial. The designer's task is not as simple as designing a product and releasing it into the world. Rather, the designer creates a landscape that constrains what successors can and cannot do in the design space. Wanton design therefore risks doing far more harm than good, particularly when attending to the distal effects of one's decisions.

One final salient, and familiar, example is the television. Modern 4k and 8k televisions are far removed from their 1927 progenitor. Despite their form and power being significantly improved, modern models are nevertheless subject to a number of constraints that result from historical choices. Institutions governing television programmes partially determine what can and cannot appear on those television sets regardless of what model someone owns. Peripheral devices (such as DVD and Blu-ray players, cassette players, and digital input devices) also play a role in determining what can and cannot be viewed and at what quality, speed and fidelity. Finally, certain homes are designed ways that constrain whether, televisions can successfully be used within their walls. As these various constraints demonstrate, it is not as simple as merely building a television. Sociotechnicity means that institutions, infrastructures, and technologies depend on each other in various ways, and that these interdependencies change over time. Given that design decisions have impacts that stretch across spatiotemporal boundaries, a close look at how these

decisions extend beyond the design space and reach out into the social world is a good place to start.

Conclusions

Technologies do not exist in isolation. They are the product of hundreds, if not thousands, of individual design decisions made across time. As a result, they become embedded in our societies, cultures, and our day-to-day interactions. Thinking carefully about how designers and engineers make their decisions is of paramount importance if we are to ensure that these technologies benefit, rather than harm, future generations.

This chapter aimed to demonstrate how trolley scenarios can be used by designers and engineers to develop ways of taking responsibility for the responsibility of others. Despite their idealized nature, traditional trolley scenarios can help us think more carefully about how real technologies can go awry. These scenarios thereby help us to conceptualize technological innovation as both a moral and anticipatory means of designing technologies in a way that avoids such vexed moral dilemmas from arising in the first place.

Bibliography

Adams, Jessica; Frankenstein, Andrea; Alabisa, James; Robinson, Tyler; Alloway, Tracy; and Lange, Lori, "Investigating the Effects of Stress on Cognitive and Emotional Moral Decision Making" (2016). Human Factors and Applied Psychology Student Conference. 3. https://commons.erau.edu/hfap/hfap-2015/posters/3

Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe?. *Philosophical studies*, *170*(3), 465-500.

Calvert, S. C., Heikoop, D. D., Mecacci, G., & van Arem, B. (2020). A human centric framework for the analysis of automated driving systems based on meaningful human control. *Theoretical Issues in Ergonomics Science*, *21*(4), 478-506.

CVL Engineers Inc. (2021, January 8). *An Engineer's Perspective on the Trolley Problem*. CVL Engineers Inc. https://cvl-eng.ca/trolley-problem/.

Fairphone. (2021, May 6). *Our Impact*. Fairphone. https://www.fairphone.com/en/impact/?ref=header.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, No. 5. Included in Foot, 1977/2002 Virtues and Vices and Other Essays in Moral Philosophy.

Friedman, B., & Hendry, D. G. (2019). Value sensitive design: Shaping technology with moral imagination. MIT Press.

Homerun.co. (2018, June 5). *Fairphone: Calling for Change*. Medium. https://interviews.artofwork.co/fairphone-calling-for-change-e4a54db6d90c.

Hotson, G., McMullen, D. P., Fifer, M. S., Johannes, M. S., Katyal, K. D., Para, M. P., ... & Crone, N. E. (2016). Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. *Journal of neural engineering*, *13*(2), 026017.

Madden, A. M. (2012). "A Revised Date for the Mosaic Pavements of the Church of the Nativity, Bethlehem". *Ancient West and East*. **11**: 147–190.

Meyer, Nicholas (Director). (1982). *Star Trek II: The Wrath of Khan* [Film]. Paramount Pictures.

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional "trolley problem". *Emotion, 12*(2), 364–370. https://doi.org/10.1037/a0025561

Nolan, Christopher (Director). (2008). *The Dark Knight* [Film]. Warner Bros. Pictures.

PERI GmbH. (2013, November 14). *Il Bosco Verticale*. Il Bosco Verticale, Milan, Italy - Projects - PERI. https://web.archive.org/web/20131207052855/http://www.peri.com/en/projects/projects/skyscrapers-towers/bosco_verticale.cfm.

Pisarchik, A. N., Maksimenko, V. A., & Hramov, A. E. (2019). From novel technology to novel applications: Comment on "An integrated brain-machine interface platform with thousands of channels" by Elon Musk and Neuralink. *Journal of medical Internet research*, *21*(10), e16356.

Rees, M. (2002, May). *The Saga of the Siege*. Time. http://content.time.com/time/subscriber/article/0,33009,1002452,00.html.

Srivatsa, N., Kaliarnta, S., & Groot Kormelink, J. (Eds.) (2017). Responsible innovation: From MOOC to book. Delft University of Technology.

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2), 204-217.

Umbrello, S., & Yampolskiy, R. V. (2021). Designing AI for explainability and verifiability: a value sensitive design approach to avoid artificial stupidity in autonomous vehicles. *International Journal of Social Robotics*, 1-10.

United Nations. (2019). Sustainable development goals. In *GAIA* (Vol. 28, Issue 2, p. 73). https://doi.org/10.14512/gaia.28.2.1

Van den Hoven, J., Lokhorst, G. J., & Van de Poel, I. (2012). Engineering and the problem of moral overload. Science and engineering ethics, 18(1), 143-155.

Van den Hoven, J., Miller, S., & Pogge, T. (2017). The design turn in applied ethics. *Designing in ethics*, 11-31.

Winner, L. (1980). Do artefacts have politics? *Daedalus*. *109*(1), 121-136.