Teaching Moral Reasoning: Why and How to Use the Trolley Problem

Johannes Himmelreich[a], Joshua Cohen[b]

[a] *Department of Public Administration and International Affairs, Syracuse University, Syracuse NY, USA*

[b] *Apple University, Apple Inc, Cupertino CA, USA; UC Berkeley, Berkeley CA, USA*

This article describes a teaching plan for a discussion-driven introduction to moral reasoning and explains its philosophical and pedagogical rationale. The teaching plan consists of a sequence of thought experiments that build on one another, and ends with participants addressing some morally complex, real-life issues. The plan rests on extensive experience teaching moral reasoning in several different professional learning environments. The main contribution of this article is practical. The goal is to equip educators with a pedagogical approach and ready-to-use teaching materials. To this end, the article offers the methodological background, identifies learning objectives as well as pitfalls of teaching the trolley problem, and describes the pedagogy of the session.

Keywords: trolley problem; teaching; moral reasoning; applied ethics

Introduction

Imagine a driver is operating an empty trolley car. As they approach a junction, the brakes fail. The driver must decide whether to follow a track leading left or one leading right. Either way, there are maintenance workers on the tracks who cannot escape in time. If the driver steers left, they will run over five maintenance workers. If they steer right, they will run over one maintenance worker. By stipulation, the six people are identical in all relevant respects, no one else is involved, and there is no option except to turn either left or right. Except for *who* dies, and for *how many* people die, the driver's actions have identical results. What should the driver do? — This scenario, and others like it, are known as the "trolley problem."

The trolley problem is perhaps the most famous thought experiment in moral philosophy (Edmonds, 2013).[1] In 2017, it was featured in the TV series *The Good Place* (cf. Meyer, 2021). In the years since, issues such as health emergency triage or the behavior of autonomous systems, which give rise to similar tragic dilemmas of life or death, have brought the trolley problem close to many people's lives (Foot, 1967; Lin, 2016; Awad et

[1] Another one is Peter Singer's analogy between aiding distant others in an emergency and saving a child drowning in a pond nearby (Singer, 1972).

al., 2018; Kamm, 2020).[2] At the same time, ethical issues in public policy and public administration have prompted broad public concern: Officeholders erode basic democratic norms, legislators are tasked to address issues of justice and fairness in the use of artificial intelligence, and civil society organizations advocate to reform police, make amends for historic injustice, and ban certain AI tools on ethical grounds.

Professional programs in public affairs need to adapt their curricula to offer education in the sensibilities and skills necessary to engage with these ethical concerns. In a recent survey published in this journal, many faculty agree that it is essential to "[help] students develop their *moral reasoning* and analytical abilities" (Baradei, 2021 our emphasis). Moral reasoning is the deliberative process of careful reflection on what one ought to do. More specifically, moral reasoning is the process of arguing for practical conclusions about what ought, all things considered, to be done in particular circumstances from premises that include general claims about value, duties, permissions, or obligations (cf. Richardson, 2018, sec. 1.1). The goal of teaching moral reasoning is neither to train ethics teachers, nor to staff ethics boards, nor to introduce students to a completely new set of ideas. Instead, the goal is to *foster a more broadly distributed appreciation of the importance of moral reflection and a more broadly distributed confidence in the ability to reason about morally complex issues—confidence that moral issues can be approached as matters of reasoning, not simply matters of taste or of gut-checking*. Public affairs programs can best achieve this goal by incorporating a practically-oriented engagement with ethical reflection into existing courses (Baradei, 2021).

This article speaks to this effort. Similar to presenting a teaching case, it aims to equip educators with a pedagogical approach and ready-to-use teaching materials that can be flexibly deployed across many courses or curricula. The article describes a session to train students in moral reasoning in an engaging and experiential fashion.[3]

However, in contrast to a teaching case, the pedagogical approach of the session needs to be grounded in normative theory. Because the session differs both in method and in substance from a usual teaching case, the article describes in detail both the method of the session as well as the concepts that students discover. Another contribution of this article is therefore that it situates the trolley problem—and related thought experiments—within the broader landscape of moral philosophy to allow teachers to conduct this session competently and confidently. The article clarifies—against widespread misconceptions—what the trolley problem is and why it is an effective way of teaching moral reasoning.

Finally, the article offers a proof of concept. It suggests that it is possible to foster classroom discussions on deeply held personal values, relate these discussions to substantive normative theories, and hone an important and urgently needed skill. The detailed

---

[2] Foot (1967) discuses a triage case: "We are about to give a patient who needs it to save his life a massive dose of a certain drug in short supply. There arrive, however, five other patients each of whom could be saved by one-fifth of that dose."

[3] A similar approach with a different sequence of cases is presented by Baltzly (2021).

description of method and substance allows teachers to make informed judgments about whether the session leads to the desired learning outcomes in their own respective learning environments.

In sum, this article contributes to public affairs teaching by describing a pedagogical approach and the associated teaching materials, grounding these materials in normative theory, and thereby providing a proof of concept. The article describes how one aspect of ethics instruction can be integrated into public affairs courses.

We will begin by describing the trolley problem and the assumptions underlying the session's pedagogy, and some potential risks of teaching the trolley problem. We then sketch a teaching plan that that addresses these risks. The plan consists of a progression of trolley problems—simple hypotheticals, designed to crystallize moral dilemmas—that introduces students to moral concepts such as consequentialism, integrity, and fairness, and conveys a sense that reasoning together about moral questions is possible.[4] The session ends with participants addressing some morally complex, real-life case(s). We suggest desiderata to identify cases that would make for a good discussion and we provide one example case. In two appendices, we reproduce the vignettes for the trolley cases and other materials that can be used to run the session.

While not an empirical case study, this article's contribution is practical as well as theoretical.[5] The article presents a ready-to-use teaching plan and provides the necessary materials; moreover, it describes the philosophical outlook that underlies its pedagogical approach. The article's methods for each of these contributions are familiar in philosophy courses. The teaching plan and materials were iteratively developed and tested in more than twenty sessions in various professional learning environments. Moreover, the teaching plan is informed by the authors' teaching experience spanning several decades and diverse audiences. So, the article presents a kind of translational research. It takes theories and teaching methods that are familiar in philosophy courses and offers a way to make them come alive in more policy-oriented settings.

The Trolley Problem

The trolley problem is not what it is often made out to be. Although it is generally presented as a dilemma, it does not consist in the dilemma itself (Foot, 1967; Thomson, 1985). And although it is called "the trolley problem", it need not involve trolleys. Instead, the trolley problem arises from considering *pairs* of dilemmas that, despite some surface similarities, elicit opposite judgments about the morally right response (Thomson, 2016, p. 115).

---

[4] Because the progression starts with the driver case and not with the bystander case, our approach stands in the tradition of what Thomson (2016) calls the "driver baptism" in contrast to "bystander baptism" version of the trolley problem.

[5] One reviewer asked us to clarify the academic contribution of this article, given that this article neither tests a theory nor develops one.

One element of such a pair is the case from the introduction in which a trolley driver has to decide whether to run over one or five workmen (Foot, 1967; Thomson, 1985). The vast majority of respondents answer that the right thing to do is to run the trolley over the one rather than over five maintenance workers.[6] And respondents are generally drawn to the idea that the action is right because it minimizes harm, or because *one death is preferable to five deaths*, all else equal (i.e. there is no other discriminating, symmetry-breaking reason or difference between the six people).

The second element to such a pair of cases is due to Thomson (1985).[7] Imagine a transplant surgeon has currently five patients, each of which is in need of some organ: One patient needs a heart, two need a kidney each, two need a lung each. The patients need these organs, or they will die very soon. One healthy person comes to the hospital for a checkup and the surgeon notices that the patient's organs would be a perfect match for the other five patients. Should the surgeon harvest the organs from the healthy person to restore the five to perfect health? The vast majority of people answer: "No".[8]

But these two judgments apparently conflict. If the trolley driver should turn right because one death is preferable to five deaths, then why should the transplant surgeon, by the same principle, *not* take the healthy person's organs?

That is the trolley problem: Given two cases that are similar in some morally salient respects but elicit apparently conflicting judgements about what to do, how can one account for the opposing judgments? Or in the words of Judith Thomson: "[W]hat difference ... explains the moral difference between [two cases]?" (1985, p. 1396). *The trolley problem consists in identifying the relevant differences that explain the apparently conflicting judgments made in response to a sequence of dilemmas*.[9] We call this the *reasoning view* of the trolley problem. According to the reasoning view, the pedagogical point of exploring the hypothetical cases is not to instruct students on ethical conduct in turning trolleys or to settle what choices to encode into autonomous vehicle software. Instead, the point is to learn how to reason carefully about moral problems, in part by formulating general principles that fit our confident judgments in more clear cases and can guide our judgment in less clear cases.

---

[6] In response to a similar case, 89% of 2,646 participants in an online experiment said that it is permissible for the trolley driver to kill one workman in order to save five (Hauser et al., 2007). However, the choice in the vignette used in this experiment was between "killing the one" and "letting the five die"—and hence not between two killings.

[7] Similar cases, of using the body of one patient to save one or more others, are discussed also by Foot (1967) and Thomson (1976).

[8] Assume that the decision is up to the surgeon, that whatever they decide will not set any precedent, that there will no repercussions or legal consequences of any form, and that the operation will restore the five patients to the same level of perfect health that the healthy person is in.

[9] The domain of dilemmas is restricted to cases that present a choice to "kill some innocent, nonthreatening person so as to either not kill or save other innocent, nonthreatening people from a threat already facing them." (Kamm, 2020, p. 82).

The trolley problem is often understood in a very different way—as being closely synonymous to "a moral dilemma about life or death" (cf. Himmelreich, 2018, pp. 671–772; Kamm, 2020, pp. 81–88). On this understanding, the trolley problem presents the question: "What do you think is the right thing for the driver to do?"[10] This can be called the *model view* of the trolley problem because it views thought experiments as representing, albeit in simplified terms, real-life situations.[11] On the model view, trolley problems represent actual dilemma situations just like models in the sciences represent phenomena of interest. Their pedagogical point would be similar to that of a flight simulator. Engaging with the problems would prepare students for unlikely but high-stakes situations, rather than aiming at developing a general cognitive skill—the capacity for moral reasoning.

The reasoning view and the model view are associated with different moral epistemologies—different theories about how humans can arrive at moral knowledge, such as whether an action is right or wrong. The model view is associated with ethical intuitionism, that is, the view that moral truths are self-evident (Stratton-Lake, 2020). By contrast, the reasoning view of the trolley problem lends itself to ethical rationalism.[12] Ethical rationalism is the view that moral truths can be arrived at by reasoning in a way analogous to how mathematical knowledge is arrived at—albeit with less stringent formalism.[13]

The teaching plan rests on the reasoning view. The pedagogical approach starts by assuming that each individual—and each student—is a competent moral reasoner. Students have a capacity to draw moral distinctions, make moral judgments, and defend those judgments with reasons. The point of engaging the cases is to improve the use of those powers by exercising them together. In addition to the fact that the reasoning view appears better suited to fill the gap of teaching the skill moral reasoning in public affairs courses

---

[10] This misunderstanding of the trolley problem was brought to great prominence not only in *The Good Place*, but also in a recent series of articles published in *Nature* and *Science*, in which scientists reported results from a large global survey on behavior in moral dilemmas (Bonnefon et al., 2016; Awad et al., 2018, 2020).

[11] Bonnefon et al. (2020, p. 112) write: "Just like the Trolley Problem and most experimental stimuli in the behavioral sciences, this autonomous car dilemma is a *model*, not a *reflection* of reality". The clarity of stating that they take the trolley problem to be a model is immediately clouded by their denial that this model does what all models are taken to be doing: reflect or represent reality in some way (Frigg & Hartmann, 2020, sec. 1).

[12] More precisely, because morally relevant differences between trolley cases are not self-evident, the reasoning view lends itself to ethical rationalism. The model view, by contrast, is typically used in a way that suggests a grounding in ethical intuitionism—with the *Moral Machine* being one example (Bonnefon et al., 2016; Awad et al., 2018, 2020). For the different good uses of trolley cases and their limitations see Himmelreich (2018).

[13] Ethical rationalism and intuitionism are philosophical outlooks that explain what moral knowledge is and how it can be obtained. The two positions hence do not rival psychological theories about how people actually behave or make decisions.

(Baradei, 2021), there are some important advantages of underwriting the pedagogical approach of this session with the reasoning view and ethical rationalism.

First, the trolley problem, as the reasoning view understands it, fosters a general skill. On the reasoning view, the trolley problem presents students with challenges that require students to pay close attention to details of cases, discern moral considerations relevant to a decision, and formulate principles that capture their judgments about cases. Once *pairs* of cases with their respective judgments are considered together and attention is drawn to the fact that the rationale offered in response to one case (Trolley Driver) conflicts with the rationale in the other case (Transplant Surgeon), students are encouraged to leave the realm of intuition, articulate the reasons for their choices, and test those reasons in subsequent cases. Moral reasoning of this kind is widely applicable, from policy analysis to legal adjudication. Teaching students to grapple with this challenge strengthens an essential intellectual skill.

Moreover, moral reasoning aims to achieve clarity about the rationales for judgments and decisions and enables a conversation among people who make different moral judgments or offer different moral reasons. In this way engaging with the trolley problem underscores that reasoning about what is right is often a social activity, in which participants are required to offer reasons that others might find compelling. Finally, and to the extent that students have good analytical skills, it brings those skills to bear on a set of problems that are sometimes thought to resist analytical thinking.

Objectives of Teaching the Trolley Problem

Presented before the backdrop of the reasoning view, the Trolley Problem can help to achieve three learning outcomes. By the end of the session outlined below, students will have a greater appreciation of:

1) **Moral reasoning**: Collaborating in reasoning together about moral dilemma situations
2) **Moral theories**: Articulating the meaning of important moral concepts and connect them to moral theories
3) **Moral dilemmas:** Reproducing the vignettes of classic thought experiments in moral philosophy and use them as heuristics to illustrate moral concepts and discuss new moral dilemmas

These three outcomes are relevant for public administration in a general and in a specific way. In a general way, attaining these learning outcomes is part of mastering critical reasoning skills. Moral reasoning requires students to draw distinctions—sometimes subtle ones—and to defend the practical relevance of those distinctions. In a more specific way, some of the trolley cases that are discussed in the session relate to case studies that present ethical dilemmas in public affairs.

Risks of Teaching the Trolley Problem

Two risks attach to teaching moral reasoning through the trolley problem.

The first risk is that students will find the thought experiments hopelessly unrealistic (Fried, 2012; Wood, 2013; Bauman et al., 2014; Kagan, 2015). The cases indeed are unrealistic: Harvesting someone's organs will not go unnoticed and will have consequences at least for the surgeon and the hospital. Moreover, it is just not possible that an organ transplantation is so successful that a patient's health is restored to precisely the level that would have achieved if they had never needed a transplant in the first place. Likewise, workmen on the two trolley tracks will differ from one another in whether they have dependents, whether they were trespassing, and in what their race, age or gender might be. — What is the point of discussing such unrealistic cases?

The second risk is that discussions of the trolley problem may foster moral relativism and cynicism about moral conduct. Faced with differing initial inclinations about what to do, participants may conclude that moral dilemmas cannot be "resolved" because they have no right answer.[14] — What is the point of discussing problems that have no solution?

If a teaching strategy fails to address these questions, students may simply—and understandably—disengage.

How to Teach the Trolley Problem

To explain our approach to gaining the benefits of using the trolley problem and forestalling these risks, we present a teaching plan for a session of 2-2.5 hours (depending on the audience and the depth of the discussion). The session is structured around a sequence of moral dilemmas, which build on another, to cover three topics: consequentialism, integrity, and fairness. Each of these topics should be seen as a kind of standpoint from which to begin one's moral reasoning.[15] As such, for pedagogical matters, the discussions of the three topics are complementary. The session ends with a discussion of one or more real cases in which students apply what they have learned (see Table 1 for an overview). In this section, we outline the pedagogical aims and the strategy for each part of the session. In two appendices we offer further materials for the session, such as the vignettes of the cases (Appendix 1), and a detailed lesson plan with definitions of concepts, quotes, definitions, and other materials that a teacher can use when facilitating the discussion (Appendix 2).

[14] Such sentiments are frequently encountered in teaching moral philosophy (Shafer-Landau, 2015, Chapter 1).

[15] Formally, each topic concerns a set of propositions of normative theory—such as necessary conditions for permissibility of an action—that form the premises of the practical reasoning.

Table 1: Overview of four parts of the session

Part	Topics	Cases
1	"Getting involved" (Consequentialism)	Trolley driver, transplant surgeon
2	"Balancing ends" (Integrity)	Captain's offer
3	"Justifying choices" (Fairness)	Bystander, Bystander with 3 options
4	Real cases	e.g. Medical triage

Motivating the Session

Because many students may not be familiar with a discussion-based format and may not have engaged with issues of ethics at length, the session begins with some motivating remarks consisting of three steps.

The first step is to introduce the general subject of ethics. This can be done by drawing attention to a widely accepted slogan that members of the class likely share, for example, that it is important "to do the (morally) right thing." The idea of "doing the right thing" is central to many organizations and institutional cultures. Alphabet, Google's holding company, admonishes employees that they "should do the right thing" (Alphabet, 2020). Similarly, the American Society for Public Administration calls on its members to "act … ethically" (ASPA, 2020). Given that such slogans are rightly afforded such a central place, it is important to share an understanding of what they mean and require.

The second step is to clarify that the session will be discussion-driven. Most students will not have a background in ethics or academic philosophy. Moreover, the topic is difficult. So expectations for a discussion-based session should be set in a way that makes students feel competent and empowered. Students should be made to realize that each of them has a capacity (albeit fallible) to reason about moral rightness. Accordingly, this session is deliberately not a lecture about morality or moral philosophy, but an exercise in *reasoning together* about morality. The idea is to convey a richer sense of what moral reasoning is by actually doing it.

Third, the theme of moral reasoning needs to be motivated and contrasted with intuitionism. What makes moral reasoning important? By reasoning we find out what the morally right thing is. By contrast, some people say that we find out what the right thing is by using our intuition, listening to an inner voice, or doing a "gut check." But with an intuitionist approach, one is unlikely to appreciate the fallibility of personal judgments or to convince others. If students want to check their intuitions, communicate *why* they think something is right, persuade others, and open themselves to criticism and self-correction, they need to reason about what is right. In short, reasoning supports clarity of thought, effective communication, and interpersonal justification.

With this three-step setup in place, the session starts its progression of moral dilemmas.

Consequentialism

This first part presents the two cases described above: the trolley driver and the transplant surgeon (see the appendix for full vignettes of all cases used in this class). The main question of this part is: What makes the moral difference between these two cases? A majority of the students will say the right thing to do for the trolley driver is to drive the trolley over the one workman, but they will not kill the one healthy patient in order to save the five patients who need organs.

The topic of this part is consequentialism (the view that the value of all actions is entirely dependent on their consequences). The aim is to help students discover, in discussion, the distinction between doing and allowing—a distinction that puts pressure on consequentialism by indicating that something matters to the moral rightness of an action other than the goodness of the results that issue from it (see appendix for a usable definition of "consequentialism"). The distinction between doing and allowing is perhaps the central difference between the trolley driver case and the case of the transplant surgeon. The trolley driver has a choice between *killing* one or *killing* five. The transplant surgeon, by contrast, faces a choice between *killing* one or *letting* five die. In other words, whereas the trolley driver has a choice between two doings, the transplant surgeon has a choice between one doing something (killing) and one allowing something to happen (letting die). To state this in a principled fashion, one would say: Although it is worse to kill five than it is to kill one, it is worse to kill one than to allow five to die (Thomson, 1985).

This part of the session starts with presenting the trolley driver case and asking students the question "What is the right thing for you to do? Why?". Students should respond by saying what they think *they ought to do*. They have to resist the temptation to predict what they would do, or to say what other people would do, or what they think other people might say, or what the law requires, or what people in "other cultures" might do or say. Three further points are worth keeping in mind.

First, teachers will want to make sure that students understand that they should imagine themselves to be *driving* the trolley. This clarification forestalls confusion of this case with a similar one that students may have encountered, called the bystander case, which makes an appearance later in this session and presents a morally very different situation. Moreover, teachers should challenge students to commit to a judgment and require that they resist the temptation to analyze the case from a third-person perspective.

Second, the case should be presented simply and without embellishment. The questions will come. Someone will ask who these workmen are and if there is any difference between them. Someone will voice the concern that the thought experiment is very unrealistic. The answers to these questions are that there is a symmetry assumption—all involved are identical in all relevant respects—and that the point of thought experiments is to offer memorable abstractions that, by isolating a few features, lead us to reason about what is morally right. We provide longer form answers as part of the lesson plan (see Appendix 2). These answers are meant to address the first risk of teaching trolley

problems outlined above, namely, that students will find the cases too unrealistic to be worthwhile. The cases *are* unrealistic—as unrealistic as balls rolling down frictionless planes. The teaching challenge is to show that they are nevertheless worthwhile.

Third, although it is important to dwell long enough on this case to get a student to articulate their reasons for running over one person—"killing five is worse than killing one" or "one should minimize the number of deaths"—it is equally important not to dwell on this case for too long. Once students have explained their reasoning, they may be quick to think about permutations of the case, such as "what if the one workman is your father?", which is usually a good sign that it is time to move on.

The next step is to present the transplant surgeon case. Again, the teacher will present the vignette and just ask the class "What is the right thing for you to do? Why?". This question is now a challenge because the answer that most students will feel is obviously right—not to harvest the one person's organs—is in tension with the reasoning the class gave on the trolley driver case, namely that "one should minimize the number of deaths".

Some students may respond by countering this challenge with questions: Who are these patients? How is this situation realistic? What is the point of discussing this? Here students are looking for more information that will break the symmetry between the cases and drive their decision. Other students may point to some purportedly obvious difference. For example, they might refer to the Hippocratic Oath or the principle of nonmaleficence to argue that a doctor must not harm people, that the healthy person has not consented, or that one ought to let nature take its course.

Both impulses—fishing for symmetry-breaking information and pointing to seemingly obvious differences—need to be resisted by asking students to elaborate. Why exactly would they not harvest the organs of the healthy person? Students will often give answers that do not stand up to reflection. They might point to professional obligations of doctors. But trolley operators, although they are not under the Hippocratic oath, also have a moral duty not to harm. Alternatively, students might point to the absence of consent from the healthy patient. But the one person on the track to the right consented as little as the healthy person in the Transplant Surgeon case. Finally, someone might offer the plausible-sounding idea that the doctor should let the five die because that is simply a matter of "letting nature take its course". But this certainly is not a rule any doctor should follow. The whole point of going to a doctor is to take nature off its course.

There are two relevant differences between the Trolley Driver and the Transplant Surgeon, and the session concentrates on one. First, the transplant surgeon uses the healthy person as a means or tool in service of the end of saving the five. This person's organs, and hence his or her death, are the *means* of saving the five who need organs. Not so in the case of the trolley driver, who steers onto the track with the one workman to avoid running into the five. The five could be saved even if the one workman were not on the other track.

The second difference between the cases is the difference between doing and allowing. As with many philosophical distinctions, this distinction is intuitive—it connects with people's sense of themselves as *agents*—but it is difficult to draw precisely (Woollard, 2015). The essential point for the class is that the distinction between doing and allowing is about *the nature of the actions themselves* and cannot easily be explicated in terms of a difference in the consequences of the actions. If students want to build their judgment on this distinction, they will have troubles with consequentialism.

The upshot of this first part, then, is threefold. First, students get a sense of "what moral reasoning looks like and feels like": it is precisely the process that they have been engaging in, of making reasons explicit and finding reasons for actions that are consistent across several cases. Second, the students have discovered a distinction between killing and letting die, or more generally, between doing and allowing. They also have the two cases, the Trolley Driver case and the Transplant Surgeon case, to make the abstract distinctions memorable. Third, the students thereby discovered a fundamental difference between two classes of moral considerations, which can be represented on the board as two columns: those based only on consequences, and those that rest on non-consequentialist considerations, for example, the nature of the action itself. The first part of the session thus advances each of the three desired learning outcomes.

Integrity: Balancing Ends

This second part of the session has three aims: (1) have students employ the considerations they just encountered, (2) introduce a case that is complex enough to draw parallels to dilemmas that students might face in their professional lives,[16] (3) allow students to realize that even a decision that is morally right might require agents to have a sense of regret.

The third case that is presented is called Captain's Offer (see Appendix 1). It is a situation of an agent who can achieve some good result only though a bad action. The protagonist of Captain's Offer case needs to either shoot one innocent person him- or herself, and subsequently four other innocent people will be let go; otherwise, if the person does not take the offer, all five will be shot by the Captain's platoon. Again, the question to students is: "What is the right thing for you to do? Why?" Classes tend to be deeply divided by this question. To many students, this case feels much harder than the two cases encountered before, although they find it hard to identify exactly what makes this case harder.

This case combines the two kinds of ethical considerations that students have identified so far. Schematically put, consequentialist considerations pull in the direction of accepting the captain's offer because one death is better than five. Non-consequentialist

[16] Although the parallels to real cases can easily be taken too far, such that distinctions, such as between doing and allowing, miss what is actually at stake (Burri, 2020).

considerations pull in the other direction: is it morally right to save the four by *killing* someone yourself? Analyzing the Captain's Offer case in this way, drawing on the distinction between consequentialist and non-consequentialist reasons, fulfills the first aim (and relates to the first two learning outcomes).

When discussing this dilemma, a student may raise the worry of relativism: People are different; they have different moral codes; each person acts on his or her own code. In response to this point, two considerations can be offered. First, an observation: It is interesting that students have not been divided about the two previous cases. On the Trolley Driver or the Transplant Surgeon, hardly anyone would react to disagreement by saying that "people are different". Indeed, worries about moral relativism rarely arise from disagreement about concrete cases. Even in the case of Captain's Offer, the problem seems to be less that different students are settled on fundamentally different views, but that each student is wrestling with a hard moral question. They may be drawn to different decisions, but they typically all feel the pull of each answer. So, whatever the merits of moral relativism may be, it seems that the cases discussed so far do not provide particularly strong evidence in its favor.

Second, even if it is assumed, for the sake of the argument, that moral relativism is correct and that the right action somehow depends on the person who is doing it, the question that the Captain's Offer case poses is: What kind of person (or organization) does *the student* want to be?

A productive way for students to approach this case is to think about it as a matter of personal *integrity*. "Integrity" is here understood, following Bernard Williams, as how someone's basic values and "ground projects" are reflected in their actions. Integrity accounts for the idea "that each of us is specially responsible for what *he* does" (Smart & Williams, 1973, p. 99; cf. Frankfurt, 1982). In Captain's Offer, an agent has to decide what is most important to him or her. Is it more important to do what has the best consequences, such as saving four lives? Or is it more important not have blood on one's hands, not be complicit with an authoritarian or unjust regime, and to respect certain "moral guardrails"?

The case of Captain's Offer serves as a focal point for students to assemble various considerations that are, on their face, non-consequentialist. These considerations are not in competition. The task that the class faces is instead a cooperative one: Each student should be asked to try to understand the respective consideration that is offered by someone, try to articulate it further, or find the right name for it—such as "complicity", "respecting basic rights", etc.

Captain's Offer offers interesting parallels to real-world dilemmas, many from public affairs. Decisions about whether or not to shoot down a hijacked plane, or whether CARE in 1994 should have provided aid to the Hutu population fleeing Rwanda (cf. Lepora & Goodin, 2013, Chapter 7), all resemble challenges presented by Captain's Offer (for more on these examples see Appendix 2).

The discussion of Captain's Offer case and associated real-life cases will lead to an important upshot: Not only are choices about integrity—conflicts between ground projects or basic values—choices of identity, of "who we are" as a person or an organization, but these choices arguably also require a kind of regret. Whatever a student chooses, they will be able to offer reasons for their choice and may think that they did what is right, all things considered. But they also have to recognize that their choice came at a moral cost (cf. Meyer, 2021). They saved four lives, but they did kill someone. Or they drew a bright line at killing someone, but five people died as a result. In short, even if someone has done the morally right thing, they often still have something to answer for. Doing the right thing and having a sense of regret might go hand-in-hand.[17]

Fairness

The third part of the session introduces the concept of fairness. Fairness shifts the moral attention from the agent to others. Whereas integrity is about the agent balancing her own ends, fairness—as we interpret it for the purposes of the session—is about justifying choices to others. In contrast to the earlier two parts, in which students "discovered" concepts in discussion, this part relies on the teacher to explain fairness.

This part begins with a presentation of the Bystander case. In the Bystander case, an empty runaway trolley is threatening five on a track ahead unless it is diverted to a sidetrack with one person. Many students will say that the right thing to do is to divert the trolley in order to save the five. But then this case seems to be a counterexample to the principle encountered in Transplant Surgeon case. The Bystander case, like the Transplant Surgeon case, presents a choice between killing one and letting five die. Reasoning about the Transplant Surgeon case led to the principle that it is worse to kill one than to let five die. Applied to the Bystander case, this principle requires likewise to let five die and not divert the trolley onto the one. The judgment that the trolley may be diverted hence conflicts with the judgments and the rationale of the Transplant Surgeon case. This is one reason for the prominence of this case (Thomson, 2016).

The concept of fairness is introduced with a story. The story is that of the late philosopher Judith Jarvis Thomson, who was a central protagonist in debates about the trolley problem—spanning publications from 1976 to 2016 (Thomson, 1976, 2016). Thomson, after several decades of thinking about the trolley problem, did something remarkable: She completely changed her mind. Whereas she had argued earlier that it is permissible to divert the trolley in the Bystander case (Thomson, 1985), she now argued that doing so is not permissible (Thomson, 2008). Why did she change her mind?

One way of putting her point is this: In concentrating on what an individual agent should do, students may have *not paid enough attention to how a decision can be justified*

---

[17] These considerations on having to admit fault usually transfer over to the real-life examples given in Appendix 2 (such as the choice CARE faced in Rwanda).

to others. This second-personal element of justifying choices to others is one way to think about fairness (see Appendix 2 for two ways of thinking about fairness). Specifically, or so argues Thomson, understanding what is permissible to do in the Bystander case requires that students consider a different case, Bystander with three options: the bystander can let the trolley continue down the track and hit five people, or divert the trolley onto a single individual, or divert the trolley onto him- or herself, thereby giving their own life to save the five.

Thomson argues in two steps. First, an agent should ask themselves if they would be willing to give their own life. She contends that one cannot be *required* to give one's life for five strangers. Giving one's life might be permissible, but it is not obligatory. Second, she argues that it is *not* permissible to divert the trolley onto the one person because doing so cannot be justified to the single individual—precisely because this person cannot be required to give their life for five strangers (Thomson, 2008, p. 367).[18]

One question for the class is whether students agree with this line of reasoning. Yet a perhaps more important question is what this concept is—of justification and fairness—that may have gone overlooked in much of the philosophical discussions on trolley problems.

The idea of justification to others is traditionally associated with contractualist theories of justice and rightness, of which Immanuel Kant, John Rawls, and Thomas Scanlon are prominent examples. But since the focus of this session is on the concepts and not on the moral theories behind them, it is more important to illustrate the idea of fairness and justification. One way of doing so, the one invoked by Thomson (2008), is to operationalize fairness in a hypothetical conversation with those who are affected by one's choices. Thus, students should imagine themselves as the Bystander engaging with the person onto whom they are going to direct the trolley. Alternatively, thinking about fairness can also be operationalized as a hypothetical convention behind a "veil of ignorance" (Rawls, 2001), that insures anonymity and thereby blocks a narrow pursuit of self-interest. With these operationalizations of fairness in mind, together with the associated trolley problems, the three learning outcomes can be achieved: that students use moral reasoning to articulate concepts—such as doing–allowing, integrity, and fairness—using the thought experiments presented in this session.

Discussion of Cases

Teaching moral reasoning has practical goals. It should foster a more broadly distributed sense of the importance of moral reflection and more broadly distributed confidence in the ability to reason about morally complex issues. To achieve these goals, real-life situations need to be engaged. Moreover—having discussed highly abstract and unrealistic

[18] Thomson (2008) suggests that Foot (1967) was right all along and that fundamentally a difference between positive and negative rights grounds the impermissibility of diverting the trolley.

thought-experiments—students should experience moral reasoning at work in real-life cases. Ending with real cases shows that the unreality of the initial cases does not keep them from being helpful in moral reasoning. Engagement with a real case hence concludes this session.

Good case studies for this part of the class have, in our experience, four features. First, the issue at hand should be hard. That is, different individuals have competing claims on some scarce good, or moral considerations pull in different directions. Second, the case should not be obscure. Ideally, the case is one with which students are already latently familiar and that is perceived to be fraught with moral issues that are hard to analyze or debate. Third, the case should be presentable with some simplifying assumptions without distorting the hard moral issue it raises. Real situations are immensely complex; any case study needs to reduce this complexity to be pedagogically appropriate. Fourth, ideally the case should be a common challenge that students face, as members of the same organization or profession.

One case that meets many of these desiderata focuses on medical triage.[19] During the Covid-19 pandemic in 2020, some hospitals had limited care resources available that were outstripped by the number of patients needing them. Chief among these were ventilators. What policy should a hospital adopt in allocating such scarce resources? This case study would put students in the role of members of a Department of Critical Care in a hospital charged with deciding on a policy to allocate scarce care resources, such as ventilators.

Although the case is realistic, it needs to be presented with some simplifying assumptions. First, the location of the hospital is not specified, thus excluding considerations pertaining to the history or the context of a particular locale. Moreover, some facts need to be stylized. For example, it is simpler to assume that anyone who is put on a ventilator will survive and leave the hospital, even if their life expectancies after leaving vary.

Both the practice of moral reasoning as well as the concepts that were identified during the session can be brought to bear on this case. One policy for allocating ventilators is consequentialist in spirit. It prioritizes those who have the longest life expectancy. This is a standard view in medical ethics, focused on years of life saved. True to the consequentialist spirit, the distinction between doing and allowing harm is assigned little relevance. Another policy, suggested by a former President of the American Public Health Association (Jones, 2020), is to allocate ventilators by a lottery. This policy may seem fair because it gives each patient the same chance of receiving care, prevents biased allocation decisions in a hospital, and avoids reinforcing socially-caused disparities in life expectancy.[20] But would such a lottery be required by the sense of fairness that was

[19] This case was developed by Joshua Cohen and Abby Jaques in a follow-on to the course that was jointly taught by the co-authors of this paper.

[20] For a further policy that has been put forward see White et al. (2020).

uncovered in the case of the Bystander? In other words: Is this the only choice of policy that can be justified to others? Such questions are bound to arise as students grapple with the question of what the right thing to do is in this case (See Appendix 2 for more details on this case).

Similar case studies abound. Potential avenues relevant to public policy and public administration include dilemmas of complicity that humanitarian workers face (Lepora & Goodin, 2013), decisions of whether to "blow the whistle" on wrongdoing—such as those faced by Daniel Ellsberg or Edward Snowden—, or decisions that city councilors need to make on regulating facial recognition. Each could be developed into a case study to practice moral reasoning and validate the concepts surfaced earlier in the session.

Concluding Remarks

Integrating ethics into an already crowded curriculum of a public affairs program is a challenge. A similar challenge is faced in many professional contexts, which emphasize the importance of moral thinking but provide only very limited time to "teach ethics." This article outlines a session that can be flexibly deployed in different courses. Engaging with the trolley problem requires moral reasoning and provides a powerful antidote to the idea that moral judgments are simply matters of personal taste, beyond the reach of reasoning. It introduces moral thinking from the bottom-up, by engaging with cases, and it fosters clear thinking and effective communication about moral issues. To be sure, some of the cases required very strained assumptions. But with the right sequence of cases—most importantly, by not beginning with the case of the Bystander—and by ending with a discussion of complex real-life moral issues, the risks involved in teaching the trolley problem can be avoided. It becomes clear in this sequence of cases that there is something to be learned from each, at least when the case is considered together with others. In a successful session, trolley cases make for a stimulating discussion, offer a way for students to articulate important moral concepts, and, crucially, connect them to real-world cases from their own experience. Most fundamentally, it conveys a sense of what it is to reason about what is right by having students actually do just that.

The Version of Record of this manuscript has been published and is available in The Journal of Public Affairs Education 10/21/21:
https://www.tandfonline.com/doi/full/10.1080/15236803.2021.1966591

Table 2: Summary of teaching materials with pedagogical suggestions and further readings

Subject	Cases	Suggestions	Additional Readings
Consequentialism	Trolley driver, Transplant surgeon	Emphasize doing/allowing distinction as well as imperative not to use others as mere means	Thomson (1985), Woollard (2012)
Integrity	Captain's offer	Emphasize interpersonal differences of ground projects as well as fact that morally permissible actions can wronging someone	Smart and Williams (1973, pp. 77–150), Cox et al. (2017)
Fairness	Bystander, Bystander with 3 options	Revisit cases and have one student who plays the role of making the choice in the case justify their choice to another student who plays the role of being one of those affected by the choice	Rawls (1971, sec. 25.1-25.2, 2001, secs. 23–25), Thomson (2008)
Real Cases	e.g. Medical triage	Select case that presents a hard ethical issue, with which students are already latently familiar, that can be presented simply, and presents a shared challenge	White et al. (2020), Jones (2020)

Appendix 1: Trolley Cases

This appendix gives materials that can be used in class. The thought experiments are derived from various places in the recent history of moral philosophy. In particular, the cases of the trolley driver, the transplant surgeon and the trolley bystander are due to Thomson ,[21] the case in which the bystander faces three options is discussed in Thomson (2008). The case of Captain's offer is a modification of what is known as "Jim and the Indians" presented by Williams (Smart & Williams, 1973, p. 98).

Trolley Driver

Suppose you are the driver of an empty trolley. You have your hands on the wheel and see five workmen on the track to the left. The track goes through a valley, and the sides are steep, so they have no way to escape. You step on the brakes, but they don't work. You can steer the trolley onto a track that leads off to the right, and thus save the five men ahead. But on this side-track there is one workman. He can no more get off the track in

[21] Although this is not the first instance of dilemmas involving the steering or directing of trolleys between groups of different sizes (Engisch, 1930, p. 288; Welzel, 1951, p. 51).

time than the five can, so you will kill him if you turn the trolley onto him.

Transplant Surgeon

Suppose you are a transplant surgeon. When you transplant organs, they take just as if you patients had been born with them and your patients are restored to full health. You have five patients who need organs. One needs a heart, two need one lung each, and two need a kidney each. If they do not get those organs, they will all die.

A healthy person whose organs are a perfect match for the five patients happens to have come in for a check-up. If you transplanted this healthy person's organs, the five patients would live. But, of course, the person is unwilling to give his life. She says, "I deeply sympathize, but no."

Captain's Offer

Suppose you are a botanist on an expedition in a politically unstable country. In the central square of a small town you find, tied up against a wall, and a row of five people and several armed men in uniform in front of them.

One man turns out to be the captain in charge. The captain explains that the five people are a random group of the town's inhabitants who, after recent acts of protest against the government, are just about to be killed to remind other possible protestors of the advantages of not protesting.

The captain offers you a "guest's privilege" of killing one of the five inhabitants yourself. If you accept, then the other four people will be let off. If you refuse, then all five will be killed. You know (from your colleagues in the anthropology department) that this offer is genuine, and the captain will keep his word.

The five men against the wall, and the other villagers, understand the situation, and are begging you to accept the Captain's offer.

Trolley Bystander

You stand by the trolley track, and you can see the situation at a glance: A trolley is out of control, hurling towards five workmen ahead. You see a switch, which you can throw, thereby turning the trolley onto the sidetrack, killing the one workman.

Bystander's Three Options

Suppose again you are a bystander. The situation is the same as before, except that there are now three options. You yourself stand on another side-track and the switch can now be thrown in two ways.

As before, if you throw the switch to the right, then the trolley will turn onto the side-track, thereby killing the one workman. But if you throw the switch to the left, then

the trolley will turn onto the side-track on which you yourself stand and you would kill yourself. Of course, you can do nothing, letting five workmen die.

Appendix 2: Lesson Plan (2 hours 15 minutes)

Part	Title	Time (minutes)
0	Introduction	10
1	Getting Involved	25
2	Balancing Ends	35
3	Justifying Choices	25
4	Discussion of cases	35
5	Summary	5
Total time		**135**

Use of the whiteboard: In first session, a teacher can collect the different considerations offered by students on a whiteboard. A two-column table structure has proven to work well. One column contains consequentialist considerations offered by students such as, maximizing welfare or number of lives saved, or quality adjusted lifeyears (QUALYs). The other column contains non-consequentialist considerations, such as the doing–allowing distinction, complicity, the rule not to kill. "Integrity", in the second part of the session, can be recommended as a process of weighing considerations within these two columns — for example by a scale above the table.

Introduction (10 minutes)

- **Motivating the session:** (1) introducing the subject (e.g. "do the right thing", "act ethically", (2) setting expectations for discussion, and (3) motivating the theme of moral reasoning.

Getting Involved, doing–allowing distinction (25 minutes)

- **Case:** *Trolley Driver*; question: "What is the right thing for you to do? Why?"
- **For questions from students**
  - **Symmetry assumption:** All individuals on the track are identical in relevant respects (whatever these respects may be). The workmen are of the same age, gender, rank, have the same number of dependents etc. We maintain this symmetry assumption for all subsequent thought experiments that we encounter today. The reason for this is that we try to isolate one consideration by keeping everything else constant.

    - **The point of thought experiments:** Thought experiments are hypothetical, extreme, and greatly idealized dilemma situations. As such, they "beg as many questions as they illuminate".[22] They beg questions because they don't tell you how you got into a situation and because they stipulate the options that are available to you. Thought experiments are *not* meant to represent real ethical dilemmas. Instead, thought experiments are meant to be memorable abstractions to investigate how we reason about what is right. They are meant to isolate certain features of the world to make them more salient. Physicists and economists build models in the same way: there is no frictionless plane and no perfect competition.
    - **The limits of thought experiments.** This is our task: Given limited information, we ask: How would you decide, and why? Because thought experiments are so abstract and artificial, they will tempt us to add detail to them by thinking "what if …" and they might lure us down various rabbit holes. This is usually a sign to move on to the next case.
- **Points to take away:**
    - **Upshot:** Killing five is worse than killing one.
    - **Consequentialism:** The right thing to do is what has the best results. You should not consider the action itself or the motives or intentions with which you act.
- **Case:** *Transplant Surgeon*; question: "What is the right thing for you to do? Why?"
- **Points to take away**:
    - **Upshot:** Killing one is worse than letting five die.
    - **Doing–allowing distinction:** Other things being equal, much stronger reasons are needed to justify doing harm than to justify merely allowing harm.
        - **Example:** Consider the difference between drowning someone and not helping someone who is drowning. You need very strong reasons to justify drowning someone (perhaps this is necessary as the only way of defending yourself). In contrast, you might be permitted to let someone drown on weaker reasons, e.g. if doing so would risk your own life. (By the way: Even if it is *permissible* to let someone drown, this does not mean that this is what you ought to do or that it is the best thing or the right thing to do in such a situation.)
    - **The trolley problem:** Reasoning by contrast. "This is what reasoning looks and feels like."
    - **Non-consequentialism:** In a slogan: "It's about the journey, not only about the destination" Consequences are not all that matters for making an action right. What also matters for reasoning about moral rightness is how a result is

[22] This is a quote from Bernard Williams.

brought about, what kind of action it is, or with what motives or intentions the action is performed.

Balancing ends, integrity (35 minutes)

- **Case:** *Captain's Offer*; question: "What is the right thing for you to do? Why?"
- **Points to take away**
  - **Integrity:** the commitments you identify with most deeply; "ground projects"
- **Similar real cases**
  - **Example:** Suppose you are a politician and terrorists have hijacked an airplane, which they will fly into a building. You can shoot down the plane, killing those in it but saving those in the building.
  - **Example:** In June 1994 Oxfam, CARE, and the International Rescue committee had to decide whether or not they want to provide aid to the mainly Hutu 2 million population that fled from Rwanda to (then) Zaire. The organizations knew that in doing so they would be supporting the Rwandan Armed Forces (*Forces Armees Rwandaises* or FAR), for example, because the FAR siphoned off aid or used the civilians as human shields. The organizations would either be "feeding perpetuators of genocide" or "denying life-saving goods services, and moral support to innocent people." Most NGOs, UN agencies chose to stay. MSF-France and the International Rescue Committee chose to leave. See also Lepora and Goodin (2013, Chapter 7)

Justifying choices, fairness (25 minutes)

- **Case:** *Trolley Bystander*; question: "What is the right thing for you to do? Why?"
- **Case:** *Bystanders' Three Options.*
- **Explanation:** Thomson contends that what is permissible for bystander to do in two-option case depends on what is permissible for them to do in three-option case.
- **Points to take away:** Thomson's (2008) argument: you cannot expect a bystander to willingly give their life to save the five (because you should also not be willing to give your own life for the five); therefore it is impermissible to impose this cost on them (even if you wanted to willingly give your life to save the five).
  - **Thomson quote:** "What the bystander does if he turns the trolley onto the one workman is to make the one workman pay the cost of his [own] good deed because he doesn't feel like paying it himself." She goes on to say: "Since he wouldn't himself pay the cost of his good deed if he could pay it, there is no way in which he can decently regard himself as entitled to make someone else pay it." (Thomson, 2008)
  - **Analogy:** Like stealing someone's wallet to give the money to charity (Thomson, 2008).

- **Fairness:** Do what others can reasonably expect of you to do.
    - **Fairness as acceptability:** A decision, and the reasoning behind it, is acceptable to everyone.
    - **Operationalization 1 (Conversation):** In a hypothetical conversation with the workman on the side-track, you would have to answer tough questions of the kind Thomson raises. The workman would ask: "Why me? Would you want to give your life? What makes you think you can decide about mine?"
    - **Operationalization 2 (Convention):** Imagine to decide on a policy of whether it is permissible to flip the switch in a hypothetical convention behind a "veil of ignorance" (Rawls, 2001), that is, each member in the convention does not know who they are in the actual situation. Now they have to decide together about the right thing to do.
- **Question:** Agree with Thomson's reasoning that fairness requires not turning the trolley in Bystander Case?

Discussion of cases (35 minutes)

Aims: (1) practice moral reasoning, (2) validate the content learned in engagement with a real-life scenario. — see article for desiderata on case studies.

Example case: Triage

- **Setup:** You are a member of the Department of Critical Care Medicine in a hospital. The Department expects to be overwhelmed with patients due to a pandemic. There will not be enough of many critical items, including ventilators. Because of the shortages, the Department forms a four-member Ethics Committee to determine how to rank candidates for ventilators. You are a member of the committee, which is asked to establish a general policy that will govern decisions, thus avoiding heart-wrenching choices in the moment. Also, the Department needs to explain treatment decisions to patients' families. The policy adopted by the Ethics Committee will provide the basis for the explanations.
- **Three positions are laid out**
    - Allocation to maximize years of life saved (a standard position in medical ethics)
    - Allocation by lottery (Jones, 2020) to avoid reinforcing socially-caused differences in life expectancy (for example, racial differences) and in hospital biases
    - A coarse-grained allocation to maximize years of life saved, subject to the condition that all lives with more than five years remaining are treated as having equivalent life expectancies (White et al., 2020)
- **Prompt:** Your group is the fourth member of the Ethics committee. You should try to answer our three questions:

1. What do you think is the right thing to do here? (Students may agree with any of the three positions, or they may offer an alternative proposal, backed by a different reason.)
2. What general rule or principle provides the best reason for that decision?
3. What do you find hardest about the case?

Summary (5 minutes)

- Introduce memorable cases
- Exemplify moral reasoning
- Surface important concepts

Anecdote from Richard Feynman on importance of knowing (and remembering) reasoning behind decisions; on his decision to join the Manhattan Project:

> "The original reason to start the project, which was that the Germans were a danger, started me off on a process of action… It was a project on which we all worked very, very hard, all co-operating together. And with any project like that you continue to work trying to get success, having decided to do it. But *what I did—immorally I would say—was to not remember the reason that I said I was doing it [for]*, so that when the reason changed, because Germany was defeated, not the singlest thought came to my mind at all about that, that that meant now that I have to reconsider why I am continuing to do this. I simply didn't think, okay?" (Feynman, 2005)

References

Alphabet. (2020, September 25). *Alphabet Code of Conduct*. Alphabet Investor Relations. https://abc.xyz/investor/other/code-of-conduct/

ASPA. (2020). *Code of Ethics*. American Society for Public Administration: Code of Ethics. https://www.aspanet.org/ASPA/Code-of-Ethics/Code-of-Ethics.aspx

Awad, E., Dsouza, S., Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2020). Crowdsourcing moral machines. *Communications of the ACM*, *63*(3), 48–55. https://doi.org/10.1145/3339904

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The Moral Machine experiment. *Nature*. https://doi.org/10.1038/s41586-018-0637-6

Baltzly, V. B. (2021). Trolleyology as First Philosophy: A Puzzle-Centered Approach to Introducing the Discipline. *Teaching Philosophy*, *44*(2).

Baradei, L. E. (2021). Ethics education in public affairs programs: What do faculty around the globe have to say? *Journal of Public Affairs Education*, *27*(2), 198–217. https://doi.org/10.1080/15236803.2020.1818023

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in

Moral Psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. https://doi.org/10.1111/spc3.12131

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2020). The Moral Psychology of AI and the Ethical Opt-Out Problem. In S. M. Liao (Ed.), *Ethics of Artificial Intelligence*. Oxford University Press.

Burri, S. (2020). Why Moral Theorizing Needs Real Cases: The Redirection of V-Weapons during the Second World War. *Journal of Political Philosophy*, *28*(2). https://doi.org/10.1111/jopp.12200

Cox, D., La Caze, M., & Levine, M. (2017). Integrity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2017/entries/integrity/

Edmonds, D. (2013). *Would You Kill the Fat Man?: The Trolley Problem and What Your Answer Tells Us about Right and Wrong*. Princeton University Press.

Engisch, K. (1930). *Untersuchungen über Vorsatz und Fahrlässigkeit im Strafrecht* (2 [1964]). Scientia Verlag.

Feynman, R. P. (2005). *The Pleasure of Finding Things Out: The Best Short Works of Richard P. Feynman*. Basic Books.

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, *5*, 5–15.

Frankfurt, H. (1982). The importance of what we care about. *Synthese*, *53*(2), 257–272. https://doi.org/10.1007/BF00484902

Fried, B. H. (2012). What Does Matter? The Case for Killing the Trolley Problem (Or Letting It Die). *The Philosophical Quarterly*, *62*(248), 505–529. https://doi.org/10.1111/j.1467-9213.2012.00061.x

Frigg, R., & Hartmann, S. (2020). Models in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2020/entries/models-science/

Hauser, M., Cushman, F., Young, L., Jin, R. K.-X., & Mikhail, J. (2007). A Dissociation Between Moral Judgments and Justifications. *Mind & Language*, *22*(1), 1–21. https://doi.org/10.1111/j.1468-0017.2006.00297.x

Himmelreich, J. (2018). Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations. *Ethical Theory and Moral Practice*, *21*(3), 669–684. https://doi.org/10.1007/s10677-018-9896-4

Jones, C. P. (2020, April 7). *Coronavirus disease discriminates. Our health care doesn't have to*. Newsweek. https://www.newsweek.com/2020/04/24/coronavirus-disease-discriminates-our-health-care-doesnt-have-opinion-1496405.html

Kagan, S. (2015). Solving the Trolley Problem. In E. Rakowski (Ed.), *The Trolley Problem Mysteries*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190247157.001.0001

Kamm, F. M. (2020). The Use and Abuse of the Trolley Problem: Self-Driving Cars, Medical Treatments, and the Distribution of Harm. In S. M. Liao (Ed.), *Ethics of Artificial Intelligence* (pp. 79–108). Oxford University Press.

Lepora, C., & Goodin, R. E. (2013). *On Complicity and Compromise*. Oxford University Press.

Lin, P. (2016). Why Ethics Matters for Autonomous Cars. In M. Maurer, B. Lenz, J. C. Gerdes, & H. Winner (Eds.), *Autonomous Driving—Technical, Legal and Social Aspects* (pp. 69–86). Springer. http://www.springer.com/de/book/9783662488454

Meyer, S. J. (2021). Everything is fine! Using "The Good Place" to teach administrative ethics. *Journal of Public Affairs Education*, *27*(2), 126–140. https://doi.org/10.1080/15236803.2020.1782102

Rawls, J. (1971). *A Theory of Justice* (1999 Revised edition). Harvard University Press.

Rawls, J. (2001). *Justice as Fairness: A Restatement*. Harvard University Press.

Richardson, H. S. (2018). Moral Reasoning. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/reasoning-moral/

Shafer-Landau, R. (2015). *The Fundamentals of Ethics*. Oxford University Press.

Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy & Public Affairs*, *1*(3), 229–243.

Smart, J. J. C., & Williams, B. (1973). *Utilitarianism: For and Against*. Cambridge University Press.

Stratton-Lake, P. (2020). Intuitionism in Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2020/entries/intuitionism-ethics/

Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, *59*(2), 204–217. https://doi.org/10.5840/monist197659224

Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal*, *94*(6), 1395. https://doi.org/10.2307/796133

Thomson, J. J. (2008). Turning the Trolley. *Philosophy & Public Affairs*, *36*(4), 359–374. https://doi.org/10.1111/j.1088-4963.2008.00144.x

Thomson, J. J. (2016). Kamm on the Trolley Problems. In E. Rakowski (Ed.), *The Trolley Problem Mysteries* (pp. 113–134). Oxford University Press.

Welzel, H. (1951). Zum Notstandsproblem. *Zeitschrift für die gesamte Strafrechtswissenschaft*, *63*, 47–56. https://doi.org/10.1515/zstw.1951.63.1.47

White, D. B., Katz, M., Luce, J., Lo, B., Daugherty Biddison, L., Toner, E., & Halpern, S. (2020). *Allocation of scarce critical care resources during a public health emergency*. University of Pittsburgh, Department of Critical Care Medicine.

https://ccm.pitt.edu/sites/default/files/UnivPittsburgh_ModelHospitalResourcePolicy_2020_04_15.pdf

Wood, A. (2013). Humanity as End in Itself. In S. Scheffler (Ed.), *On What Matters* (Vol. 2, pp. 58–82). Oxford University Press.

Woollard, F. (2012). The Doctrine of Doing and Allowing I: Analysis of the Doing/Allowing Distinction. *Philosophy Compass*, *7*(7), 448–458. https://doi.org/10.1111/j.1747-9991.2012.00491.x

Woollard, F. (2015). *Doing and Allowing Harm*. Oxford University Press.