Journal of Ethics and Social Philosophy
VOL. 30, NO. 3 · MAY 2025
https://doi.org/10.26556/jesp.v30i3.4324

IN DEFENSE OF THE TROLLEY METHOD

Ian Stoner and Jason Swartwood

GUY CRAIN recently argued that the trolley method of moral philosophy has three shortcomings that have not yet been appreciated by its practitioners.[1] To the extent that Crain's criticisms highlight ways fanciful examples are sometimes abused, we welcome them. We are moved to reply because Crain suggests that he has identified shortcomings of fanciful examples as they are routinely employed by philosophers. We disagree.

1. WHAT FANCIFUL EXAMPLES ARE FOR

Our fundamental objection to Crain's approach is that his characterization of fanciful examples in moral philosophy neglects one of their key features: fanciful examples, competently employed as philosophical tools of persuasion or critical reflection, always have the goal of shaping the beliefs of readers in some specific way.[2]

Fanciful examples play a variety of belief-shaping roles. They ground arguments from analogy, function as paradigm cases for abducting moral principles, and present scenarios that invite conceptual clarification. But perhaps the most common use of fanciful examples (and of described cases in general) occurs within the method of wide reflective equilibrium, where they "function as either *counter-examples* or *reductio ad abdsurda*."[3] Authors develop examples intended to elicit from readers a reflectively endorsed moral judgment about a concrete case that is clearly in tension with a specific target belief. Used this way, fanciful examples are a "custom-built tool for illuminating the ill-fittingness of a target belief."[4]

1 Crain, "Three Shortcomings of the Trolley Method of Moral Philosophy," (hereafter cited parenthetically).

2 We prefer the umbrella term 'fanciful examples' for any wholly invented, heavily stipulated described case. The metonymic 'trolley case' is needlessly misleading; Crain's and others' criticisms of this method cover many examples that do not feature trolleys.

3 Walsh, "A Moderate Defence of the Use of Thought Experiments in Applied Ethics," 471.

4 Stoner and Swartwood, "Fanciful Examples," 326.

In this discussion note, we focus on fanciful examples used as counterexamples. The well-known cases Crain catalogs in the introduction to his article are all arguably deployed as counterexamples to specific target beliefs. Consider three of the fanciful examples he highlights.

Organ Transplant targets the belief that it is morally permissible to kill one person in order to save five people.[5] Judith Jarvis Thomson's audience is readers who are tempted by the original trolley problem to abstract the belief that the lives of the many outweigh the life of one. Thomson expects that such readers will judge that it is *not* morally permissible for a surgeon to kill one healthy person in order to transplant his organs into five desperate patients, and this judgment is obviously in tension with the belief that it is permissible to kill a person in order to save many others.

Ticking Time Bomb targets the belief that torture is always, in principle, wrong.[6] Those who have used this example expect that readers will form the judgment that torture is permissible in situations in which it is the only way to save many innocent lives, and this belief is in tension with the belief that torture is categorically impermissible.

Jim the Botanist targets the belief that ethical thinking is a simple algorithm that takes as input all and only those levers of power currently available to the agent.[7] Bernard Williams expects readers will find it hard to form a judgment about what Jim should do when Pedro offers to spare the lives of nineteen condemned innocents if Jim personally executes one of the twenty captives. Even though there is only one lever of power available to Jim, and the consequences of pulling it are clear, the decision Jim faces is not trivially easy to settle.

Attempts to deploy fanciful examples as counterexamples fail when they fail to elicit a reflectively endorsed judgment about the described case that is in tension with the target belief. There are many available paths to such failure. Some examples dictate stipulations that are difficult or even impossible for most readers to imagine.[8] Some examples invite unreliable judgments that arise from framing effects or bigoted background beliefs.[9] Failures such as these are mistakes in applying the method, not weaknesses of the method itself.

Crain's target is the method itself. Some of Crain's objections are prudential objections, others are methodological. Our position is that Crain's objections

5 Thomson, "Killing, Letting Die, and the Trolley Problem," 206. Thomson's Footbridge example (207–8) is structured the same way and targets the same belief.

6 Alhoff, "Ticking Time-Bombs and Torture."

7 Smart and Williams, *Utilitarianism*, 98–99.

8 McGee and Foster, *Intuitively Rational*, 161–63.

9 Wood, "Humanity as an End in Itself," 69; and Stoner and Swartwood, "Fanciful Examples," 330–33.

all fail, for the underlying reason that he has neglected the belief-shaping goal of fanciful examples. Once we acknowledge that examples aim to shape beliefs in specific ways, it is clear that the objections Crain develops are not objections to the use of fanciful examples but objections to their abuse.

2. Methodological Objection: Trolley Cases Fail to Model Real-World Ethical Decisions

In the real world, most ethical decisions are influenced by past experiences and feature limited conscious reflection; real-world decisions usually occur under some degree of uncertainty and rarely involve discrete decision points with enumerated options. Real ethical decisions, in other words, are nothing like trolley problems. Crain suggests that this failure to model the conditions of real-world decisions undermines the trolley method as a method of moral philosophy (433–34).[10]

The trouble with this objection is that it misrepresents the methodological role of fanciful examples. They are not intended to model ordinary moral decision-making. They are intended to test specific moral beliefs that an audience holds using that audience's own judgments about described cases. The goal of examples is to improve the set of moral beliefs that we will, in turn, employ in real-world decision-making.

Consider an analogy to the difference between medical research and everyday medical practice. Imagine a randomized controlled trial designed to test a new cancer drug. That trial will yield new information about the effectiveness of the drug, and that information will in turn shape the treatment plans oncologists develop for their patients. The researchers' study of drug efficacy is, in an important way, involved in practical decision-making about treatment, but only when clinicians apply in the practical domain the information generated by the research study. Different methods are appropriate for determining the general efficacy of the drug and deciding whether the drug is appropriate for a particular patient in a particular situation, and it would be absurd for a clinician to object that randomized controlled trials are methodologically flawed on the grounds that they fail to model the complicated, fluid decisions clinicians make in partnership with their patients in choosing treatment plans.

Similarly, philosophers who use fanciful examples aim to improve the set of beliefs we employ in practice. Just as with medical research and medical practice, these are two different areas of inquiry. There is no reason to expect that

10 See also Wood, "Humanity as an End in Itself," 70; and Fried, "What Does Matter?" 506.

philosophical methods that refine a set of beliefs should mirror the application of that set of beliefs in practical decision-making.

Take, for example, Peter Unger's trolley case, Bob's Bugatti.[11] Unger's target is the belief that charity's demands must be limited—people who make significant sacrifices on behalf of strangers may deserve praise, but their sacrifices are supererogatory. Against this belief, Unger sketches a situation in which a man invests his retirement savings in a Bugatti. He parks his retirement Bugatti just beyond a railroad siding and goes for a hike. On returning, he spots a child trapped on the main track in the path of an oncoming train. Bob can direct the train onto the siding, which will destroy his retirement Bugatti, or he can do nothing, in which case his retirement investment will remain secure, but the child will be crushed. Bob chooses to let the child die. In response to this case, nearly everyone forms the judgment that Bob's decision is monstrous.[12]

Bob's Bugatti does not model typical real-world decisions; Crain is right that we rarely or never find ourselves facing choices like Bob's. But Unger has targeted a belief that is involved in familiar practical decisions about charitable giving. Someone who believes that morality cannot demand much in the way of charity will think differently about how to set up their monthly budget than someone who does not hold that belief. Just as a randomized controlled trial can help a clinician improve their beliefs about the efficacy of the treatments they prescribe, Bob's Bugatti can help us improve our beliefs about the limits of our charitable obligations.

When careless thinkers attempt to treat fanciful examples as models for real-world choices, disaster sometimes ensues. This has been hashed out at length in discussions of Ticking Time Bomb. Ticking Time Bomb, properly employed, can demonstrate that hardly anyone believes, on reflection, that torture is always, in principle, wrong. But it is a gross abuse of that case to argue that because torture is the right thing to do in a wildly unrealistic fantasia, torture should be an option available to real-world governments.[13] People who have defended real-world torture by appeal to Ticking Time Bomb are reasoning poorly *because* they are treating the fanciful example as a model of real-world decision-making when, like most fanciful examples, it is not and should not be.

11 Unger, *Living High and Letting Die*, 136.

12 In addition to functioning as a counterexample, Bob's Bugatti can play other belief-shaping roles, such as functioning as a paradigm case for an inference to the best explanation (which is one of the roles it plays in Unger's original) or as the basis of an argument from analogy (which is how Peter Singer uses the same case in "The Singer Solution to World Poverty").

13 Beck and de Wijze, "Interrogating the 'Ticking Bomb Scenario.'"

3. Methodological Objection: Because They Involve High Stakes, Trolley Cases Are Useless for the Ethics of the Mundane

Fanciful examples are typically extreme and often violent; they involve death and dismemberment, earthquakes and boat accidents, puppy murders and disintegrator rays. The questions most of us face in daily life are nothing like that. Should I call my mom? Should I microwave my fish sticks in the employee break room? May I pass along a juicy morsel of gossip, or must I leave it unplucked? This gulf between the enormity of fanciful examples and the mundanity of the ethical choices most of us actually face leads Crain to object that "the trolley method is, by design, a terrible tool for working on the ethics of the mundane" (427).

If it were true that the extreme features of fanciful examples make them useless for mundane ethics, then that would be a mark against the method. But in many familiar domains—perhaps most famously, physics and mathematics—extreme cases, carefully employed, can help focus and correct the beliefs and patterns of reasoning that we apply to realistic cases.

Take, for example, a classic puzzle of mathematics, the Monty Hall problem. Monty Hall presents you with three doors, only one of which hides a valuable prize. You make your selection—say, door three. Monty Hall then opens one of the other doors—say, door one—revealing that there is nothing behind it. He then gives you the opportunity to change your guess. Should you stick with door three or change your guess to door two?

When first presented with the Monty Hall problem, many people's statistical judgments are mistaken—they believe that the odds are even that they have picked the winning door, and switching their guess would be pointless. One way to make clear that this belief is false is to imagine an extreme version of the case. Suppose Monty Hall presents you with one hundred doors, and behind one of them is a prize. You guess door three. He then opens ninety-eight of the remaining doors, revealing nothing behind them—only door three and door two remain closed. Should you stick with door three or switch your guess to door two? In this case, it is obvious that you should switch your guess. Recognizing this makes it easy to see that in the real-world, three-door version of the game, you should likewise switch your guess. The extreme case helps us achieve a clarity of understanding that helps us navigate the initially murkier real-world case.

Fanciful examples in moral philosophy can play a similar role. When it comes to our obligations of charity to strangers, some of our moral beliefs are arguably mistaken. Bob's Bugatti is an extreme case intended to help. It shines a light on a target belief that many of us have adopted from our families and communities: that significant charitable giving is supererogatory. By providing

us with a fanciful example that defamiliarizes charitable giving, Unger makes it easy for us to see that Bob's choice to prioritize his own secure retirement is not, by our own lights, morally permissible; significant levels of financial sacrifice *can* be obligatory. Recognizing this makes it easier to see that the mundane question of how to structure our monthly budget is morally fraught.

Bob's Bugatti is not an outlier. Other fanciful examples are structurally similar. They often target beliefs that are centrally implicated in the ethics of the mundane. They often accomplish their goal by providing extreme cases that allow us to see those beliefs in a clearer, because less familiar, light.[14] Fanciful examples are obviously not the only tool for working on everyday moral problems, but they can help improve beliefs relevant to the ethics of the mundane. The fact that extreme cases in any domain—physics, mathematics, and ethics, too—can be misleading when used carelessly is a reason to use them carefully, not to reject them categorically.

4. METHODOLOGICAL OBJECTION: FAUX ANONYMIZATION OF AGENTS INVITES DISTORTED RESPONSES

Crain notes that in most fanciful examples, "but for the sparsest of features, the agent is anonymized" either via minimally sketched third-person cases or second-person cases in which each reader is explicitly invited to imagine themselves as the agent (430). Crain suggests that anonymization is problematic because readers supply their own details for these minimally described cases, and regardless of their backgrounds, readers tend to supply details that track closely with Western, educated, industrialized, rich, and democratic (WEIRD) cultural assumptions (429). The result of faux anonymization is that non-WEIRD agents "are likely being operationally excluded.... The trolley method is not ethics for everyone" (431).

Crain's suggestion that non-WEIRD readers supply WEIRD details for anonymized agents runs counter to our own classroom experiences. But suppose he is right. Suppose readers, regardless of their own backgrounds, tend to imagine WEIRD agents in anonymized cases—most everyone, and not just our mothers, picture George Clooney when invited to imagine Thomson's transplant surgeon. This would not undermine the value of Organ Transplant as a counterexample to the target belief that it is morally permissible to kill one to save many. Anyone who imagines George Clooney in Organ Transplant still

14 Walsh, "A Moderate Defence of the Use of Thought Experiments in Applied Ethics," 473–74.

has before them a concrete case in which they themselves judge that it would be wrong to kill one to save many.

The same is true for other fanciful examples employed in their various belief-shaping roles. Examples can be effective in their persuasive roles even if non-WEIRD readers imaginatively supply WEIRD cultural details, so long as those readers respond to the example in the way the author intends. If WEIRD beliefs prompt reactions to the case that are morally misleading or incorrect, that merely provides additional beliefs the method can target for revision, not a reason to reject the method.

In our classrooms, we have noticed that students, including students from non-WEIRD communities, often supply details drawn from their own experiences when presented with under-described cases written by other students. It sometimes happens that the culturally specific details they supply flip the valence of the judgment the case's author intended to elicit. That is a marker of a case that is *problematically under-described*.[15] An effective counterexample must be broadly accessible and uncontroversial.[16] If anonymization of agents typically results in examples that are so under-described that they are unable to accomplish their persuasive role, then that would constitute a methodological criticism of anonymization. But there is no reason to think that anonymization of agents typically risks problematic under-description of a case. Note that in every one of Crain's opening examples, including Organ Transplant, Jim the Botanist, and Ticking Time Bomb, agents are anonymized, and yet these are famous examples in part *because* most readers, regardless of their backgrounds, form in response to them the judgments that their authors intend to elicit.[17]

In most cases, anonymization of agents is a kindness authors extend to their readers. Anonymization makes it easier for us all to imagine *ourselves* in the situation the author has sketched, with all of our existing values, beliefs, backgrounds, and experiences. Consider a de-anonymized version of Thomson's Organ Transplant:

> *Sara the Transplant Surgeon*: Sara is a thirty-seven-year-old divorced mother of two. She was raised in a conservative Muslim home, and

15 This is the case in Crain's own Bystander case (429). That deliberately under-described case lacks sufficient detail for readers to judge whether intervention is warranted.

16 Stoner and Swartwood, *Doing Practical Ethics*, ch. 3.

17 To be clear, eliciting a common judgment from a diverse audience is not sufficient for these or any other examples to succeed as counterexamples. The elicited judgment must withstand reflective scrutiny and be in genuine tension with the target belief. It might turn out, after further argument, that everyone's judgment about the case is misguided and should be revised. We simply intend to highlight that anonymization of agents does not undermine the effectiveness of otherwise well-designed cases.

> though she still observes some practices, she has quietly come to think of herself as agnostic. Her relationship with her parents was strained by her divorce, but over the previous few months they have made efforts to extend olive branches to each other. Sara finds her work as a surgeon rewarding, but she is buffeted by regrets: of her huge outstanding student loans, of her curtailed time with her children, of her status as perpetually on call. She worries that she bonds with her patients more than other surgeons do and that she may be substituting connection to her patients for her attenuated connections to her parents and children. Should Sara murder one innocent bystander in order to save five of her patients?

Sara the Transplant Surgeon reads as a gentle parody because it is clear that her biography is irrelevant to Thomson's project. What matters to Thomson is that *you*, reader, with *your* beliefs, *your* values, *your* biography, judge that it would be wrong to kill an innocent in order to harvest their organs.[18] Part of what makes Thomson's example effective is that readers' judgments about it are not contingent on their biographical idiosyncrasies. WEIRD and non-WEIRD alike, most readers agree that harvesting organs is no excuse for murder.

5. PRUDENTIAL OBJECTIONS

In addition to his methodological objections, Crain worries about the impression fanciful examples leave on students and the public. Fanciful examples, with their characteristically high stakes "build the impression that ethics just amounts to the rare decision faced by many or the normal decision faced by the few" (427). Relatedly, since it is lower-stakes decisions that "make up the stuff of everyday moral life" (426), fanciful examples contribute to the impression that philosophers are uninterested in the ethical issues most people face. When authors anonymize agents, that opens the possibility that they will "insidiously create the impression that to be a moral agent at all is to be a WEIRD, young, and fit male" (432). These prudential objections persist even if we are right that Crain's methodological objections fail. But we wish to highlight that prudential objections provide no reason to avoid the method of fanciful examples in philosophical research.

18 We thank an anonymous reviewer for pointing out that the best characterization of Thomson's method remains an area of interpretative debate. See, e.g., Conte, "The Trolley Problem and Intuitional Evidence." Our argument does not rest on whether Thomson would agree that we have correctly characterized her method. We expect she would agree that her anonymization of the surgeon in Organ Transplant does not undermine the effectiveness of the example.

Philosophy teachers and public philosophers should recognize that philosophical methods, including the method of fanciful examples, are unfamiliar to many people. Audiences unfamiliar with the trolley method—whether students or members of the public—should be supplied with the background they need to understand its means and goals. Poor communication of the trolley method, in classrooms and in public venues, risks doing harm.[19]

But just as the analogous challenges of science education and science communication do not give scientists reason to abandon scientific methods, the challenges of philosophical education and public philosophy do not give us reason to abandon philosophical methods. We should do our best to teach our students and to educate the public about the role fanciful examples play in refining the set of moral beliefs that shape our everyday moral life. If the method is useful, as we have argued it is, then the risk of public misunderstanding gives us reason to work harder at teaching well, not reason to abandon the method.

6. CONCLUSION

We have argued that Crain's objections to the trolley method of moral philosophy can be defused by taking seriously the belief-shaping role of fanciful examples. We have focused on fanciful examples as counterexamples to specific target beliefs, but our arguments can easily be adapted for fanciful examples used in other specific belief-shaping roles: as analogies, paradigm cases, and so on. Competently employed in pursuit of the goal of shaping beliefs in these ways, fanciful examples can be extraordinarily useful philosophical tools.

Although Crain's objections fail as criticisms of standard uses of the trolley method, his discussion serves to highlight an important way examples can fail. When authors, teachers, and public figures fail to identify the specific beliefs their fanciful examples are designed to shape, they flirt with several follies.[20]

Saint Paul College
ian.stoner@saintpaul.edu
jason.swartwood@saintpaul.edu

19 Martena, "Thinking Inside the Box."

20 We are grateful to three anonymous reviewers for comments that improved this discussion note.

REFERENCES

Allhoff, Fritz. "Ticking Time-Bombs and Torture." In *Contemporary Debates in Applied Ethics*, edited by Andrew I. Cohen and Christopher Heath Wellman. 2nd ed. John Wiley and Sons, 2013.

Beck, Simon, and Stephen de Wijze. "Interrogating the 'Ticking Bomb Scenario': Reassessing the Thought Experiment." *International Journal of Applied Philosophy* 29, no. 1 (2015): 53–70.

Conte, Sebastian J. "The Trolley Problem and Intuitional Evidence." *Journal of Value Inquiry* 58, no. 2 (2024): 293–310.

Crain, Guy. "Three Shortcomings of the Trolley Method of Moral Philosophy." *Journal of Ethics and Social Philosophy* 26, no. 2 (2023): 420–43.

Fried, Barbara. "What Does Matter? The Case for Killing the Trolley Problem (or Letting It Die)." *Philosophical Quarterly* 62, no. 248 (2012): 505–29.

Martena, Laura. "Thinking Inside the Box: Concerns About Trolley Problems in the Ethics Classroom." *Teaching Philosophy* 41, no. 4 (2018): 381–406.

McGee, Andrew, and Charles Foster. *Intuitively Rational: How We Think and How We Should*. Springer Nature, 2024.

Singer, Peter. "The Singer Solution to World Poverty." *New York Times Magazine*, September 5, 1999.

Smart, J. J. C., and Bernard Williams. *Utilitarianism: For and Against*. Cambridge University Press, 1973.

Stoner, Ian, and Jason Swartwood. *Doing Practical Ethics*. Oxford University Press, 2021.

———. "Fanciful Examples." *Metaphilosophy* 48, no. 3 (2017): 325–44.

Thomson, Judith Jarvis. "Killing, Letting Die, and the Trolley Problem," *Monist* 59, no. 2 (1976): 204–17.

Unger, Peter. *Living High and Letting Die*. Oxford University Press, 1996.

Walsh, Adrian. "A Moderate Defence of the Use of Thought Experiments in Applied Ethics." *Ethical Theory and Moral Practice* 14, no. 4 (2011): 467–81.

Wood, Allen. "Humanity as an End in Itself." In Derek Parfit, *On What Matters,* vol. 2. Oxford University Press, 2011.