JAHR | Vol. 15/2 | No. 30 | 2024

Original scientific article / Izvorni znanstveni članak
Received / Zaprimljen: 24. 10. 2024.

https://doi.org/10.21860/j.15.2.9

Samuel Kahn*

Kantian Trolleyology

SUMMARY

This article examines six contemporary Kantian approaches to Judith Jarvis Thomson's famous trolley problem, a thought experiment that continues to challenge philosophers and non-philosophers alike. The approaches analyzed include those of Paul Guyer, Pauline Kleingeld, Samuel Kerstein, Elke Elisabeth Schmidt, James Edwin Mahon, and Allen Wood. All of these philosophers attempt to reconcile Kant's moral framework on the one side with the ethical issues raised by the trolley problem on the other. However, this article highlights and elaborates on significant challenges to each philosopher's approach, pointing to potential flaws and inconsistencies. In engaging critically with these varied Kantian approaches, the article not only highlights their limitations, but also revisits Thomson's original framing of the problem in order to assess the latter's enduring significance and to delve into some of the systemic issues inherent in trolleyology. It is argued that recognizing these systemic issues is crucial for advancing a more robust Kantian solution to the trolley problem, one that remains faithful to Kantian principles while nonetheless addressing the complexities of moral decision-making in such situations.

Keywords: trolleyology, Kantian trolleyology, Kant's ethics, Kantian ethics, Thomson, duties.

SECTION 1. THOMSON'S TROLLEY

Thomson's original trolley problem is deceptively simple. Consider the following three thought experiments:

> **Trolley**. A trolley is hurtling down a track, and if it continues, it will run over five people who are on the track and have no means of escape. The trolley cannot be stopped, but you could throw a switch that would cause the trolley to change onto an alternate track. If the trolley changes onto the alternate track, it will run over one person who is on that track and has no means of escape. The question is: Should/may you throw the switch? (Kahn, 2023, pp. 487–488)[1].

* *Correspondence Address*: Samuel Kahn, Department of Philosophy, Indiana University, Indianapolis, 425 University Blvd, Indianapolis, IN, 46205 USA. E-mail: kahnsa@iu.edu. ORCID: https://orcid.org/0000-0001-6597-7646.

[1] Thomson's original articulations of the trolley problem are to be found in her 1976 and 1985. For context, see (Kahn, 2023, nn. 1–3) and (Mahon, 2021, pp. 156–162, and esp. 156n8).

> **Fat Man**. A trolley is hurtling down a track, and if it continues, it will run over five people who are on the track and have no means of escape. You are on a bridge that goes over the track, and there is a fat man standing next to you. By shoving the man, you could cause him to fall down onto the track. He would die in the process, but his body would prevent the train from running over the five. The question is: Should/may you shove the fat man? (Kahn, 2023, pp. 487–488).

> **Loop**. A trolley is hurtling down a track, and if it continues, it will run over five people who are on the track and have no means of escape. The trolley cannot be stopped, but you could throw a switch that would cause the trolley to change onto an alternate track. The alternate track curves back around to the five. But if the trolley changes onto the alternate track, it will run over one person who is on that track and has no means of escape, and the person's body will prevent the track from completing the loop and running over the five. The question is: Should/may you throw the switch? (Kahn, 2023, pp. 487–488).

According to Thomson, Loop is variant of Trolley, and in both cases, it is intuitively permissible to save the five. But, in Fat Man, it is intuitively impermissible to save the five. The challenge is to come up with a principle that can explain all three of these cases (without being falsified by any other).

Kantian ethics enters the picture because one of Kant's principles famously tells us never to use humanity merely as a means. This prohibition, sometimes known as the Kantian prohibition, seems promising if we confine our attention to Trolley and Fat Man: the person on the alternate track in Trolley is not used as a mere means to save the five, whereas the fat man in Fat Man is.

But, according to Thomson, if we now expand our scope to Loop, we encounter a hitch. This is because "there is no plausible account of what is involved in, or what is necessary for, the application of the notions "treating a person as a means only," or "using one to save five," under which [the agent in Fat Man] would be doing this whereas the agent in this [Loop] variant of [Trolley] would not be" (Thomson, 1985, p. 1403). Thomson concludes that "these notions [of treating persons as means or using one to save five] cannot do the work being required of them here" (Thomson, 1985, p. 1403). Thus, according to Thomson – and her argument "is widely considered convincing" (Kleingeld, 2020b, p. 204)[2] – the Kantian prohibition founders on Loop.

[2] I should note that I am not able to find any evidence to corroborate this assertion of Kleingeld's. In the cross-section of the massive trolleyology literature with which I am familiar, Thomson's argument against the Kantian prohibition seems more widely to be ignored than endorsed, and Kantians seem to pick it up only in order to explain where it goes wrong.

SECTION 2. GUYER'S APPLICATION OF KANT BY THE NUMBERS

Guyer maintains that the Kantian prohibition, interpreted correctly, can be used to solve the trolley problem. His argument has two steps.

First, Guyer argues for a novel interpretation of the Kantian prohibition. Guyer points out that, according to Kant's *Lectures on Ethics*, it might have been permissible for the Roman leader Cato to commit suicide if his goal was to encourage the Romans to defend their freedom more vigorously (Guyer, 2006, p. 197). On the basis of this, Guyer suggests that "the (freely chosen) destruction of one free being in order to save many more free beings may be permissible, or even mandatory, because making humanity in both our own person and that of all others an end and never merely a means might well require preserving as many instances of humanity as possible" (Guyer, 2006, p. 197). That is, Guyer advocates interpreting the Kantian prohibition as having a quantitative aspect. According to Guyer, a prohibition on using people as mere means requires saving the greatest number of people, even when that, in turn, requires sacrificing some for the purpose.

Second, Guyer applies this interpretation to a variation of Trolley, arguing that "it is not merely permissible but even obligatory for you to throw the switch so that only one person is killed by the train" (Guyer, 2006, p. 198). Guyer reasons that the Kantian prohibition requires us to save the five by sacrificing the one, as per Thomson's intuition. Although Guyer does not consider Loop, his reasoning can be generalized to apply to it as well. Thus, Guyer's application of Kant-by-the-numbers serves to explain Thomson's intuition that it is permissible, or perhaps even obligatory, to save the five in Trolley and Loop.

However, there are at least three problems with Guyer's approach.

The first problem is historical. Kant's lecture notes are not a good source for substantive interpretation of his thought. For one thing, these notes were not written by Kant. They were written by his students, and they were written in prose form after, not during, the lectures. That is, in addition to being removed from the source, these notes are extrapolated. For another thing, it is difficult to know when a lecturer is saying something that they sincerely, reflectively, and consistently endorse, rather than trying out a new idea or blundering through a muddled one. So, in general, any interpretation of Kant that is based primarily on these lecture notes is highly speculative.

The second problem is exegetical. Guyer's interpretation of the Cato example is underdetermined. An alternate interpretation of the passage is that suicide is permissible in this instance because to continue living would make Cato complicit

in the violation of the Kantian prohibition[3]. The notes equally can be interpreted as suggesting that suicide was permissible for Cato because, if he continued to live, then he would be captured and tortured and, thence, used as an example (as a mere means) to compel the Romans to submit (27: 370). Thus, although the student notes can be interpreted as Guyer advocates, they also can be interpreted otherwise.

The third problem is philosophical. If Guyer's reasoning is accepted, then, all else being equal, it is permissible, and perhaps even obligatory, to save the five in Fat Man. That is, if the Kantian prohibition permits us, and perhaps even obliges us, always to preserve as many instances of humanity as possible, then, all else being equal, the Kantian prohibition permits us, and perhaps even obliges us, to push the fat man off the bridge and into the path of the trolley, using him as a mere means to save the five. This is contrary to Thomson's intuition. So, Guyer's interpretation of the Kantian prohibition fails to explain the trolley problem with which we began.

Now, Guyer might maintain that all else is not equal in Fat Man. That is, Guyer might argue that, on his interpretation of the Kantian prohibition, we are required to preserve as many instances of humanity as possible only if doing so does not require us to use anybody as a mere means. Thus, Guyer might reason that, because we are using the fat man as a mere means in Fat Man, his interpretation does not imply that we are permitted to save the five in this case—in fact, quite the alternative. On these grounds, Guyer might conclude that he has solved the trolley problem after all.

However, this line of reasoning does not work. Recall that, in Loop, saving the five also seems to require using the one as a mere means. That is precisely why Thomson originally argued that the Kantian prohibition founders on the trolley problem. This puts Guyer in a bind. It looks like he cannot explain Fat Man without giving up on Loop, and vice versa, exactly as Thomson contends. If we want a solution to the trolley problem, we must look elsewhere.

[3] Mahon also objects to Guyer's interpretation of the text. However, Mahon's objection is different from mine:

> Kant nowhere permits an exception to the perfect duty not to commit suicide. The casuistical questions concerning suicide in *The Metaphysics of Morals*, as well as in his lectures on ethics, are about whether certain acts are acts of suicide, and not whether acts of suicide are permissible (Mahon, 2021, p. 181).

According to Mahon, Kant thinks that the prohibition on suicide is exceptionless, and texts that suggest otherwise actually should be interpreted as questioning whether a given action is an instance of suicide, not whether a given action is an instance of permissible suicide.

I disagree with Mahon. But, I should note that this is a verbal dispute. Nothing substantive stands or falls on whether we understand the case of Cato as a permissible instance of non-suicidal killing oneself, or as a permissible instance of suicide: Guyer's interpretation can weather Mahon's objection.

SECTION 3. KLEINGELD'S COMPROMISE SOLUTION

Kleingeld advocates an alternate solution to the trolley problem. Her solution has three steps.

First, Kleingeld advocates interpreting the Kantian prohibition in terms of agents' practical reasoning: "An agent uses another person as a means if and only if she wants to reach a certain end, believes that she can reach or further this end mediately, by using another person as a means, and uses the person for the sake of reaching or furthering her end" (Kleingeld, 2020b, p. 212). This is sometimes referred to as an agent-centered rather than a patient-centered approach (Kahn, 2024a, section 1). The idea is that, although there might be an external description of an agent's actions according to which she is using someone as a mere means, she does not count as violating the Kantian prohibition unless this description matches her intentions in some way. For example, Cha might sit down on Bart when she needs a rest, using Bart as a conveniently placed seat. But, if this is because Bart is lying in a sleeping bag and looks (to Cha's myopic eyes) like a log, then Cha is not violating the Kantian prohibition, at least as Kleingeld advocates interpreting it (Kleingeld, 2020a, pp. 399-400).

Second, Kleingeld argues that the Kantian prohibition should be interpreted in terms of consent (Kleingeld, 2020a; 2020b). More specifically, according to Kleingeld, whether an agent uses another as a mere means depends on whether her use of the other is contingent on the other's actual consent. For example, suppose that Cha knows that it is Bart, rather than a log, lying in front of her. If Cha asks Bart for his permission to sit down before doing so, and if he consents (or if, at any rate, she thinks that he consents), then Cha is using Bart as a means, but she is not using Bart as a mere means on Kleingeld's account. If, by way of contrast, Cha could not care less whether Bart consents, then her post-perambulatory perch is, in fact, a violation of the Kantian prohibition (even if Bart chips in to say that he is OK with it).

Third, Kleingeld applies her agent-centered, consent-based interpretation of the Kantian prohibition to the trolley problem. Kleingeld argues that in Trolley, an agent who saves the five is not using the one as a means, and, *a fortiori*, such an agent is not using the one as a mere means (Kleingeld, 2020b, p. 215). So, if it is impermissible to save the five in Trolley, Kleingeld argues, this will not be on account of the Kantian prohibition[4]. In Fat Man, by way of contrast, Kleingeld maintains that an agent who pushes the fat man "necessarily uses him merely as a means" (Kleingeld, 2020b,

[4] Kleingeld is circumspect in her conclusion because, as she points out, the Kantian prohibition is a proper part of the Categorical Imperative, and, considered in full, the Categorical Imperative might make saving the five in Trolley impermissible, even if the Kantian prohibition does not (Kleingeld, 2020b, p. 226).

p. 217). This, according to Kleingeld, is because, in Fat Man, the trolley can be stopped *only* by using the fat man merely as a means, whence it may be concluded that any agent who stops the trolley must be doing so (Kleingeld, 2020b, p. 218). Turning, finally, to Loop, Kleingeld maintains that "we can ascribe different lines of practical reasoning to the agent, such that the agent does or does not use the heavy workman merely as a means" (Kleingeld, 2020b, p. 218). Kleingeld argues that Loop is structurally similar both to Trolley and to Fat Man. Thus, it is possible to imagine an agent in Loop violating the Kantian prohibition by acting on the same maxim as an agent in Fat Man (namely: "I will save more rather than fewer human lives, even if this involves my using others as means to this end without their actual consent") (Kleingeld, 2020b, pp. 218-219). However, it also is possible to imagine an agent in Loop acting in accordance with the Kantian prohibition by adopting the same maxim as an agent in Trolley (namely: "I will save more rather than fewer human lives, provided I do not use anyone as a means to this end without their actual consent") (Kleingeld, 2020b, pp. 219-220). In this way, Kleingeld argues that the Kantian prohibition is consistent with saving the five in Trolley and in Loop, but it is inconsistent with saving the five in Fat Man and in Loop. This, of course, is revisionary: in the original trolley problem, saving the five in Loop is permissible; Thomson does not consider that Loop might allow for alternate construals. But, Kleingeld maintains that this revisionary aspect of her account is one of its strengths, because it enables her to explain "why the [Loop] scenario has generated such radically different moral assessments in the literature" (Kleingeld, 2020b, p. 222).

However, there are at least two problems with Kleingeld's account.

The first problem is exegetical. In the *Metaphysics of Morals*, which contains Kant's most extended and detailed taxonomy of duties (with corresponding derivations), Kant makes repeated use of the Kantian prohibition. Therefore, if consent-based interpretations of the Kantian prohibition were correct, one would expect to find this reflected in the *Metaphysics of Morals*—an expectation that is dashed upon inspection of the text. This militates against Kleingeld's consent-based interpretation of the Kantian prohibition[5].

The second problem is philosophical. Kleingeld's claim that an agent in Loop can save the five without using the one as a mere means is, as she herself admits, based on "what we stipulate" (Kleingeld, 2020b, p. 220). But, stipulation is not open to Kleingeld at this juncture of the dialectic. As we have seen, Thomson argues, quite convincingly by Kleingeld's own admission, that an agent is using the one as a mere means to save the five in Fat Man if but only if she is doing so in Loop. So, merely

[5] Kleingeld's consent-based interpretation has come under heavy fire in the secondary literature (Seymour Fahmy, 2021, pp. 5–6 and 2023; Kahn, 2024a).

stipulating that an agent can save the five without using the one as a mere means in Loop—stipulating that it is open to the agent to adopt the maxim "[to save the five] provided I do not use anyone as a means to this end without their actual consent" - is question-begging[6].

Now, Kleingeld might reply that I am being uncharitable. For one thing, what she intends to stipulate is not the question-begging claim that the agent in Loop is able to adopt a maxim not to use the one as a mere means. Rather, Kleingeld is stipulating the details of the case such that it is rational for an agent in Loop to believe and desire in such a way as to make this maxim psychologically within reach. For another thing, Kleingeld might point out that the maxim in question is not about the unanalyzed notion of using someone as a mere means; rather, it is about the conjunction of (1) using someone as a means without (2) making this use contingent on their consent. Thus, even if Kleingeld were merely stipulating that such a maxim is rationally available to an agent in Loop, this stipulation would not be question-begging because this conjunctive maxim is not the same as the one that Thomson rules out.

However, this reply does not help. Kleingeld's rationale for asserting that any agent in Fat Man who saves the five violates the Kantian prohibition applies also to any agent in Loop who saves the five. To see why this is so, we must note two things. First, Kleingeld asserts that, in all three cases, Trolley, Fat Man, and Loop, the agents do not get the actual consent of the one. Thus, on Kleingeld's interpretation of the Kantian prohibition, if the agents in any of Trolley, Fat Man, or Loop use the one as a means to save the five, then the agents use the one as a mere means to save the five. This is precisely why Kleingeld concludes that "[t]he crucial question in each of the three trolley scenarios, therefore, is whether the agent *uses* the one man as a *means*" (Kleingeld, 2020b, p. 215). Second, as seen above, the agent in Trolley, on Kleingeld's account, does not violate the Kantian prohibition because she is not using the one for anything. But, according to Kleingeld, the agent in Fat Man does violate the Kantian prohibition because "the action of pushing the heavy man off the bridge (given the details of the case) necessarily involves the agent's using him merely as a means" (Kleingeld, 2020b, p. 225). And the problem, which is hopefully now clear, is that this explanation of Fat Man can be picked up and applied to Loop: exactly as in Fat Man, there is no way for the agent in Loop to save the five without killing the one as a means to stop the trolley, whence it might be concluded that it is impossible for an agent in Loop to save the five without violating the Kantian prohibition. Kleingeld's attempt to introduce a morally relevant distinction between Fat Man and Loop does not withstand critical scrutiny; if we are going to save the Kantian prohibition from the trolley problem, this switch is not the one to pull.

[6] Versions of this objection may be found in (Kahn, 2023, section 2) and (Schmidt, 2022, pp. 205–206).

SECTION 4. KERSTEIN'S REVISIONARY APPROACH

Kerstein takes a more revisionary approach than Guyer or Kleingeld.

Kerstein advances three alternate interpretations of the Kantian prohibition. Two of the interpretations are consent-based; the third is about whether the patient of an action can contain, in herself, the end of that action. The details of these interpretations are unimportant for present purposes. What is important is that all three are agent-centered. The way that this manifests is that the operative questions for Kerstein are, first, whether, in Trolley or in Loop, an agent who proposes to save the five is using the one to that end and, if so, then, second, whether, in Trolley or in Loop, an agent who proposes to save the five reasonably can believe either (first interpretation) that the one can stop them by dissenting to their action, or (second interpretation) that the one would consent to their action, or (third interpretation) that the one can share their end (Kerstein, 2013, pp. 123-124).

Kerstein, like Kleingeld, maintains that, in Trolley, an agent who proposes to save the five "would not be *using* the person or persons [they] do not save" and, *a fortiori*, such an agent is not using anyone as a mere means (Kerstein, 2013, p. 123). Thus, Kerstein concludes that, in Trolley, the Kantian prohibition has "no implications" (Kerstein, 2013, p. 122)[7].

However, Kerstein maintains that things are otherwise in Loop, and this is where the revisionary nature of his account enters in. In Loop, according to Kerstein, an agent who proposes to save the five would be using the one. Moreover, because it is not reasonable for such an agent to believe any of the three things set out above, Kerstein argues that, in Loop, saving the five violates the Kantian prohibition on all of the three alternate interpretations he offers (Kerstein, 2013, p. 124). In the face of Thomson's intuition that it is permissible to save the five in Loop if, but only if, it is permissible to save the five in Trolley, Kerstein maintains otherwise, asserting that "the moral distinction between the two cases...is the fact that in Loop...you treat the one person merely as a means, but in Trolley you do no such thing" (Kerstein, 2013, p. 124).

However, in parallel with Kleingeld's account, there are at least two problems with Kerstein's account.

The first problem is exegetical. The objection made above against Kleingeld's consent-based account applies equally to Kerstein's two consent-based accounts, and it now can be extended to cover Kerstein's end-sharing account: just as we do not

[7] Like Kleingeld, Kerstein leaves open the possibility that saving the five might be impermissible for reasons other than the Kantian prohibition (see note 4, and the sentence to which it is appended, above).

find consent playing a major role in the *Metaphysics of Morals*, so we do not find end-sharing playing a major role in the *Metaphysics of Morals*, exactly the opposite of what we would expect if Kerstein's third interpretation was correct. So, all three of Kerstein's interpretations of the Kantian prohibition appear to be on shaky exegetical ground.

The second problem is philosophical. Thomson admits that, in Loop, saving the five involves using the one as a mere means and, thus, that saving the five in Loop runs afoul of the Kantian prohibition. Indeed, as seen in section 1 of this paper, the point of the trolley problem is precisely to explain why it is permissible to save the five in Loop despite the fact that doing so violates the Kantian prohibition. So, as Kleingeld points out, Kerstein's line of reasoning "is unlikely to satisfy those who, like Thomson herself in her 1985 paper and again in her most recent discussion of the issue (2016, pp. 128–132), are wondering just *how* the addition of a bit of track [in Loop] causes such a massive moral difference" (Kleingeld, 2020, p. 210)[8].

SECTION 5. MAHON'S AND SCHMIDT'S KANTIAN RIGORISM

Mahon and Schmidt independently advocate an even more revisionary position than Kerstein's: they argue that Kantian ethics, correctly interpreted, implies that it is impermissible to save the five in all of the trolley cases.

Schmidt's main argument for this position builds on the thesis that trolley cases should be interpreted as a conflict between a narrow duty not to kill the one and a wide duty to save the five (Schmidt, 2022, section 3). Because, according to Schmidt, narrow duties always trump wide ones on Kant's account, a duty not to kill the one trumps a duty to save the five and, therefore, it is impermissible to save the five in all of the trolley cases. Schmidt appeals to Kant's notorious murderer at the door case in order to defend her interpretation (Schmidt, 2022, pp. 209-210). As Schmidt reads this case, it involves the conflict of the narrow duty not to lie and the wide duty to save a life; the narrow duty not to lie always trumps the wide duty to save a life in

[8] It is no help to Kerstein that, in her 2008, Thomson contends that her original intuition was mistaken and that it is never permissible to turn the trolley: the shift from Thomson's early work to her late work is in whether she thinks an agent may sacrifice the one to save the five; Thomson remains committed throughout to the biconditional that it is permissible to sacrifice the one to save the five in Trolley if, but only if, it is permissible to do so in Loop. Thus, regardless of whether Kerstein takes on the early or the late version of the trolley problem, his attempt to introduce a morally relevant distinction between Trolley and Loop faces an uphill battle, at least against Thomson. This is not to say that Kerstein cannot emerge victorious from such a battle. But, it is to say that, if he is engaging in trolleyology, he at least must enter this battle.

Kantian ethics; and, just so, the narrow duty not to kill also always trumps the wide duty to save a life[9].

Mahon independently makes the same argument:

> This case, therefore, involves a conflict between a negative (perfect) legal duty not to kill one innocent person (not to commit murder), and a positive imperfect ethical duty to save five innocent people from being killed (or to not let five people die). According to Kant's moral theory, the negative (perfect) legal duty is more stringent than the positive imperfect ethical duty; or rather, there is no conflict, since imperfect duties have perfect duties 'built in' them as exceptions (Mahon, 2021, p. 183)[10].

However, Mahon supplements this argument with another. Considering the case from the perspective of rights, Mahon argues that to kill the one would involve an infringement of the right to life, whereas to let the five die would not:

> The case, therefore, involves throwing a railroad switch and diverting a runaway train to a sidetrack, so that it kills just one person, which infringes a right of that one person, and not throwing the railroad switch, and allowing the runaway train to continue on its track and kill five people, which infringes the right of no-one. It is clear that Kant would not uphold infringing a right of someone and doing someone a wrong. Hence, throwing a railroad switch and diverting a runaway train to a sidetrack so that it kills just one person is prohibited by Kant's moral theory if the case is considered in terms of rights (Mahon, 2021, pp. 185–186).

Thus, on Mahon's account of Kant, there are two reasons to think that saving the five is impermissible in all of the trolley cases: (1) saving the five would violate a narrow duty in favor of a wide one, even though narrow duties always trump wide ones, and (2) saving the five would involve a rights violation, and rights violations trump imperfect duties.

This, of course, puts both Schmidt and Mahon at odds with Thomson's intuitions in the trolley problem. But, Schmidt and Mahon handle these intuitions very differently (despite the similarities in their interpretations of Kantian ethics).

Schmidt, at one point in her article, suggests that the clash between intuitions and Kantian ethics is simply so much the worse for the latter: "Even if we might not like this result from a systematic, non-Kantian ethical point of view, it is the result Kantian ethics leads to" (Schmidt, 2022, p. 198). This seems to point toward the

[9] The inference in this sentence, from the duty not to lie to the duty not to kill, is complicated: as Schmidt points out, Kant is in favor of capital punishment, and Kant also thinks that killing in cases of necessity is unpunishable (Schmidt, 2022, pp. 210–211). However, these complications do not concern us here.

[10] Mahon also, like Schmidt, appeals to Kant's murderer at the door to defend this argument (Mahon, 2021, pp. 183–184).

rejection of Kantian ethics. However, in her conclusion, Schmidt is less conciliatory toward these conflicting intuitions:

> [A]lthough intuitions cannot (and should not) be banned completely from practical philosophy, we should not rely on them blindly. Rather, we should be ready to dismiss some of them, and this is exactly what a Kantian analysis of the trolley problem brings home (Schmidt, 2022, p. 218).

Thus, Schmidt's considered position is that we should be open to dismissing our intuitions about trolley problems if they disagree with Kantian ethics.

Mahon takes a different approach. Whereas Schmidt is skeptical of the use of intuitions in practical philosophy, Mahon is not—or, at least, he does not express any such skepticism in his discussion of the trolley problem. Instead, Mahon relies on the fact that, in subsequent work, Thomson rejects her original trolley problem intuitions, arguing, instead, that it is never permissible to save the five in the trolley cases (Thomson, 2008). In explaining Thomson's later intuitions from a Kantian perspective, Mahon takes his work to be done, notwithstanding her earlier intuitions.

The main problem for both Schmidt's and Mahon's arguments, however, is that neither of them deals with the Kantian prohibition directly. Regardless of whether we take Thomson's early intuitions or her late intuitions as the benchmark, and, in fact, regardless of whether we reject the use of intuitions in practical philosophy altogether, the point of the trolley problem, at least as originally formulated and at least for Kantian ethics, is to examine how the Kantian prohibition handles these cases. So, in explaining how the narrow/wide duty distinction handles the trolley problem, and in explaining how a rights-based approach handles the trolley problem, Schmidt and Mahon have jointly, if independently, gotten onto the wrong track, even if their approaches can be given a Kantian pedigree: they have, in effect, sacrificed the Kantian prohibition on the altar of Kantian ethics, even though the latter was never brought into question (except through the Kantian prohibition).

Now, Schmidt and Mahon might maintain that I am being unfair. They might contend that both the narrow/wide duty distinction and Kantian rights are derivable from the Categorical Imperative in general and from the Kantian prohibition in particular. Thus, they might argue that, if there is a solution to the trolley problem that appeals to the narrow/wide duty distinction or to rights violations, and if the narrow/wide duty distinction and Kantian rights are derivable from the Kantian prohibition, then this solution can be reframed in terms of the Kantian prohibition directly. On these grounds, they might conclude that my objection only scratches the surface of their approaches; once we dig deeper, we can see that there is a ready reply.

However, this reply does not withstand critical scrutiny. The problem is that it is far from clear that the narrow/wide duty distinction and Kantian rights, when derived from the Categorical Imperative, work in the way that Schmidt and Mahon need them to. To see this, we can look at Schmidt's and Mahon's actual attempts to ground their approaches on the Categorical Imperative.

Mahon models the derivation of the narrow duty not to kill on Kant's derivation of the narrow duty not to commit suicide in the *Groundwork to a Metaphysics of Morals* (GMS, AA 04: 421-422). Mahon reasons that suicide violates the Kantian prohibition because it involves using one's own person as a mere means toward avoiding an intolerable condition. Thus, Mahon argues, correlatively, murder violates the Kantian prohibition because it involves elevating the maintenance of a tolerable condition over the victim's humanity:

> The ethical argument against murder would be that the surgeon, in cutting up the healthy person and distributing his organs (without his consent) to those who will die without them, is elevating maintaining lives in which people are in at least a tolerable condition above the humanity of the person who is killed to ensure this end (Mahon, 2021, p. 182).

However, this does not work. On the one side, the surgeon in Mahon's case would be committing murder even if the beneficiaries of his action would not be in a tolerable condition after receiving the planned transplants; and on the other side, an agent in the trolley cases need not sacrifice the one in order to maintain the five in a tolerable condition—it could be viewed simply by the numbers, as per Guyer's interpretation, as a sacrifice of one life for five, rather than in terms of life as opposed to pleasure. Following these two points to their respective logical conclusions, we may see that, on the one side, Mahon fails to provide a sustainable derivation of the duty not to murder, and, on the other, the derivation he provides fails to explain why, on his account, it is impermissible to save the five in the trolley cases.

Schmidt is no better off. Schmidt concedes that "*no one* is used as a mere means in the original trolley case, regardless of whether the switch is flipped or not" (Schmidt, 2022, p. 212). In other words, in Trolley, neither letting the five die, nor saving the five and killing the one, violates the Kantian prohibition. Accordingly, in order to hold on to her thesis that Kantian ethics implies that it is impermissible to save the five and kill the one in Trolley, Schmidt appeals to the full Formula of Humanity, the formulation of the Categorical Imperative from which the Kantian prohibition is taken: "So act that you use humanity, whether in your own person or the person of any other, always at the same time as an end, never merely as a means" (GMS, AA 04: 429, emphasis omitted). With this in hand, Schmidt argues as follows:

> [I]t is not only not permissible to use someone as a means, but it is also not permissible to carry out actions that do not do justice to a person's unconditional worth, that is, to his or her dignity. In this (second) sense, someone is also used as a mere means when he is not treated with the appropriate respect he deserves...Flipping the switch, however, means violating the duty to respect the dignity of the one, for not to kill is a narrow duty (Schmidt, 2022, pp. 212–213).

However, there are three distinct problems here. First, Schmidt mistakenly assimilates the failure to use someone at the same time as an end with using someone as a mere means. But, Kant does not think that these are the same, and, in the *Metaphysics of Morals*, he even gives an example of how an agent can fail to use someone at the same time as an end without using that person as a mere means (namely: by adopting a maxim of indifference toward that person) (MS, AA 06: 395). So, Schmidt's assimilation of the failure to use someone at the same time as an end with using someone merely as a means is exegetically flawed[11].

Second, Schmidt's claim that, in Trolley, saving the five does not violate the Kantian prohibition even though it involves killing the one is fatal to her overall position. As Kant explains his taxonomy, narrow duties, like the duty against suicide and the duty against lying promises, flow from the Kantian prohibition, whereas wide duties, like the duty of self-improvement and the duty of benevolence, flow from the requirement to use others always at the same time as ends. Thus, if flipping the switch to save the five in Trolley does not violate the Kantian prohibition, as Schmidt claims, then her attempt to interpret the trolley problem as involving a conflict of a narrow duty with a wide one is sunk.

Third, Schmidt's claim, in the final sentence of the block quotation above, that flipping the switch violates the narrow duty not to kill and *therefore* violates the Categorical Imperative, is question-begging in this context.[12] The explanation at this point of the argument needs to go precisely in the other direction: Schmidt needs

[11] This point, about the distinction between using someone as a mere means and failing to use someone at the same time as an end, is discussed in greater depth in Kahn (2024a, n8).

[12] Similarly, in her discussion of the universalizability formulations of the Categorical Imperative, Schmidt argues as follows:

> There appears to be a contradiction between killing and the general prohibition of killing innocent persons as described above (which would be circular)... As we worked out above, to kill innocent persons is forbidden, according to Kant; but that is one thing to say. Another thing to say would be to point to a specific contradiction, where this contradiction is supposed to be the reason for the prohibition in the first place (Schmidt, 2022, p. 212).

Schmidt admits in this passage that (1) to make a bald appeal to the duty not to kill innocent persons is fallacious (it "would be circular"), and (2) she is unable to "point to a specific contradiction" that can show how this duty can be derived from the universalizability formulations. On the basis of this, Schmidt tries to distance herself from the universalizability formulations, asserting that they are "burdened with difficulties" (Schmidt, 2022, p. 212). But this distancing, coupled with the fact that Schmidt is forced to do the same with the Kantian prohibition, undermines the Kantian pedigree she claims for her approach.

to explain how her discussion of the narrow duty not to kill can be traced back to the Kantian prohibition, not the other way around, and, in fact, if the two points I already have made are correct, then this is precisely what she cannot do, for she already has admitted that the duty not to kill the one in the trolley cases cannot be derived from the Kantian prohibition.

I want to make one last point about Mahone and Schmidt. One of the hallmarks of trolleyology is that, every time a principle is proposed to explain a set of cases, a variant case is proposed that runs this principle to the ground (Friedman, 2002). Unfortunately, that approach threatens to apply here as well. Suppose we modify all three of the original trolley cases in the following way: the five who are on the track have been put there by you and, in fact, you also set the runaway trolley in motion—and now you are faced with the same choices as in the original cases, whether to divert the trolley onto the track with the one, or whether to shove the fat man off the bridge. Call this the trolley-problem-2.0. We might want to recalibrate our intuitions before trying to find a principle to explain trolley-problem-2.0. But, even if we do so, it seems unlikely that the approaches offered by Mahone and Schmidt will shed much light on the results[13].

SECTION 6. WOOD'S REJECTION OF TROLLEYOLOGY

The revisionary nature of these approaches peaks in Wood, who rejects trolleyology almost wholesale.[14]

Many of Wood's criticisms of trolleyology are independent of Kantian ethics. For example, consider the following representative passage:

> "[T]rolley problems" often abstract artificially from the fact that it would surely be illegal for a mere bystander to touch the switches on a trolley. Or alternatively, they stipulate matters that would in any real situation be quite uncertain, such as whether farther down the track on which you see one person standing, there might be a dozen others just out of sight...Even if there is consensus about a given problem, it is seldom clear what moral beliefs the consensus response might be registering, especially where the examples involve artificial assumptions and abstract from facts about what we would and would not know in real life. If this is unclear, we should not

[13] In fairness to Schmidt, her primary goal is not to engage in trolleyology; rather, she engages with the trolley problem as a means, in order to advance toward questions about the ethics of autonomous driving (Schmidt, 2022, p. 192).

[14] The one concession Wood makes to trolleyology is that, on account of human vulnerability and wickedness, it is likely that there are always going to be situations in which we must make "stark trade-offs between the deepest interests of different people and groups" in the way that trolley problems depict (Wood, 2011, p. 91). I owe this reference to Kleingeld (2020b, 205n3).

> regard responses to these examples as credible data for moral epistemology (Wood, 2008, p. 50)[15].

In this excerpt, Wood objects to trolley problems on the grounds that they artificially abstract from relevant information and artificially stipulate away uncertainty.

However, these objections do not have to do with Kantian ethics *per se*, nor does Wood present them otherwise. Indeed, such complaints have been taken up by nonKantians, such as Fried, who shares Wood's negative assessment of trolleyology:

> Allen Wood devotes a substantial portion of his commentary in *On What Matters* to decrying the outsized role of trolley problems in nonconsequentialist philosophical argument. I share many of his objections, including to the fantastical nature of the dilemmas trolley problems pose; the absence of contextual information that in real life changes the moral complexion of tragic choices; and the unrealistic stipulation that the outcomes of all available choices are known with certainty ex ante (Fried, 2012, p. 509)[16].

> Of the various moral principles that have emerged from the now four-decades-long preoccupation with trolley problems, none can handle the problem of garden-variety risk. As a result, trolleyology is at best engaged in what amounts to a moral sideshow. (Fried, 2012, p. 506).

It is also notable that some of these objections might be raised against Kant himself. For instance, Kant's murderer at the door example is arguably as cartoonish as the trolley problem—as Wood is well aware, having done so much exegetical work to explain how the artificiality of this example obscures the ethical issues with which Kant was grappling (Wood, 2008, chapter 14).

Precisely because these objections are not distinctively Kantian, they fall outside the purview of the current investigation.[17] However, Wood does raise an objection to trolleyology that is, if not uniquely Kantian, at least one that probably would be associated mainly with Kantian ethics.[18]

According to Wood, trolleyology is based on a view of ethics and moral epistemology that "grounds principles on the consilience of intuitions" (Wood, 2008, p. 46). Wood then points out, by way of contrast, that Kant grounds his moral principles in the faculty of reason: these principles are known *a priori*, independently of experience,

---

15 This criticism is repeated in Wood's 2011 (esp. pp. 70, 82, and 82n26).

16 I owe the reference to Fried to Kleingeld (2020b, 205n3).

17 It is perhaps worth noting, however, that, although many of Wood's criticisms seem well justified, he somewhat dulls their rhetorical force when he himself takes a firm stand on what the agents in some of these trolley problems ought to do (see, e.g., his remarks about the case he refers to as "Lifeboat" in his 2011, p. 71).

18 I would like to thank an anonymous reviewer for pointing out to me that intuitionists might make a similar objection.

and if we come up against cases that elicit intuitions that seem to contradict them, then it is always and ever so much the worse for the intuitions (Wood, 2008, chapter 3; see also Wood, 2011, part one). This turns the trolley problem (and most trolleyology) on its head: if, as Thomson maintains, the Kantian prohibition cannot explain our intuitions, then either this is because we have not sufficiently scoured the logical universe of possible explanations, or it is because something has gone awry with the intuitions themselves.

I want to say two things about this. First, as a matter of exegesis, I think Wood gets Kant right: Kant makes it clear, in the *Groundwork to a Metaphysics of Morals* and elsewhere, that the moral law is *a priori* synthetic, not subject to revision based on intuitions elicited from thought experiments. This does not mean that Kant never appeals to intuitions. For example, at least some of Kant's claims about the unconditioned goodness of a good will, in part I of the *Groundwork to a Metaphysics of Morals*, seem to be based on intuition—Kant puts them forward as the considered judgments of an impartial rational spectator, underived from a higher principle.[19] But, the key point here has to do with the epistemic status that these judgments are supposed to have: they are, again, *a priori* synthetic, and, as such, not subject to revision—there is no reflective equilibrium of the kind that seems to be baked in to much trolleyology.

However, I think that at least some modern philosophers, Kantians and nonKantians alike, will find this dissatisfying because they reject Kant's epistemological model.[20] Of course, Wood might argue that this rejection is incoherent. But, that is contentious, and any such argument is beyond the present scope. Moreover, even those who accept Kant's epistemological model might point out that there is room for doubt regarding whether Kant has alighted upon the correct formulation of the *a priori* synthetic moral law, and, it might be argued, consideration of the trolley problem can serve as a healthy check upon too dogmatic assertion to the contrary.

Second, even those who accept both Kant's epistemological model and Kant's formulation of the *a priori* synthetic moral law in general, as well as the Kantian prohibition in particular, might find a use for trolleyology, notwithstanding Wood's Kantian critique. This is because, as will become important below, Kant's ethics is about more than merely the moral law and how this law applies to act tokens; it is also, as Kant makes clear in the *Metaphysics of Morals*, about general duties, which function as what Mill might have called secondary principles and what we might call auxiliaries:

[19] I would like to thank an anonymous reviewer for reminding me of this.

[20] For example, Kerstein is quite explicit about his espousal of an alternate epistemological model (Kerstein, 2013, section 1.2, remarked upon in Kahn, 2014)—as, of course, are the foils Wood mentions in chapter 3 of his (2008).

intermediary principles that guide us in day-to-day life, or at least describe it, every bit as much as the moral law, if not more so—and the point for present purposes is that trolley problems might impinge on these auxiliaries rather than the moral law. With that in mind, I want to return, briefly, to the trolley problem before advancing my own account, or at least the direction that I think an account should take.

SECTION 7. THOMSON'S TROLLEY RECONSIDERED

Recall that, in the original trolley problem, there are three cases, Trolley, Fat Man, and Loop, and Thomson asserts that it is permissible to save the five in Trolley and Loop but impermissible to save the five in Fat Man. On the basis of this, Thomson asserts that the Kantian prohibition is unable to explain Loop, and from this, she infers that the Kantian prohibition is unable to explain Trolley and Fat Man. (Or, correlatively, in her later work, when, as noted above, Thomson flips her intuitions on Trolley and Loop, the inference can move from the inability of the Kantian prohibition to explain Trolley to its inability to explain Loop and Fat Man.)

As may be seen from the foregoing, Kantian responses to the trolley problem focus on Thomson's intuitions about the cases: they argue that, *pace* Thomson, these intuitions can be explained by appeal to the Kantian prohibition (Guyer, 2006); that the intuitions are incomplete—and, once completed, they can be explained by the Kantian prohibition (Kleingeld, 2020b); that the intuitions about Loop should be discarded (Kerstein, 2013); that the intuitions about Trolley and Loop should be discarded (Mahon, 2021; Schmidt, 2022); or that all the intuitions are suspect and any intuitions that disagree with the Kantian prohibition should be discarded (Wood, 2011). However, all of these Kantian responses leave untouched Thomson's inference from the failure of the Kantian prohibition in Loop to the failure of the Kantian prohibition in Trolley and Fat Man. And it is precisely this inference that I want to call into question now.

There are three reasons for being suspicious of this inference.

First, it is at least questionable whether the failure of a principle to explain one case entails anything about its failure to explain other cases, even when those other cases are in the vicinity of the original ones. This is an issue more frequently discussed in the philosophy of science than in ethics, but perhaps the philosophy of science discussion should be extended. If Newtonian mechanics or the ideal gas law retain their explanatory power, as, to judge from the way that science is taught around the world, they do, despite their broader failures and idealizing assumptions, then, all else being equal, we might think that ethical principles can do the same.

Second, concentrating on this inference makes more apparent, I think, that Thomson's Loop is a garden-variety attempt at a false negative for the Kantian prohibition. The reason I think this is important is that it lends more weight to revisionary approaches that call into question Thomson's intuitions. That is, even if we thoroughly reject the Kantian epistemology highlighted by Wood and embrace a non-foundationalist approach to ethics, a principle as explanatorily powerful, and with as much historical and cultural momentum, as the Kantian prohibition surely ought not to be discarded merely on the basis of a single intuition about a highly unrealistic and artificial thought experiment.

Third, many of the alternate trolley cases that cause intuitions to switch differ in form rather than in substance. In this, trolleyology is much like Williams' famous personal identity thought experiments (Williams, 1970). Let me explain. Williams proposes two thought experiments involving memory erasure, one of which is supposed to elicit intuitions in favor of psychological continuity theories of personal identity, the other of which is supposed to elicit intuitions in favor of physiological continuity theories of personal identity. The philosophical punch, then, is not merely that we have contradictory intuitions, but, more, that these intuitions are elicited from alternate descriptions of the same scenario. Along the same lines, the three cases in the original trolley problem are consistent with the three cases in trolley-problem-2.0 from section 5 of this paper, and Fat Man is consistent with what we might call Fat-Man-3.0, in which, 10 minutes prior to the fateful trolley coming into sight, the fat man makes you promise to toss him in front of a trolley should the opportunity present itself for him thereby to save the five people on the track ahead. Indeed, Trolley and Loop can be formulated as elaborations of the same scenario. But, how can adding information change either whether an action is permissible or whether someone is used as a mere means?

This, I want to suggest, should give pause—which is exactly what I think the Kantian approach recommends. Let me explain.

SECTION 8. THE LAST SECTION

I want to distinguish between two questions:

1. Is it consistent with the Kantian prohibition for a particular agent in a trolley case to save the five?
2. Is it consistent with the Kantian prohibition, in general, for agents in trolley cases to save the five?

These are not the only questions that could be asked about the trolley problem cases, and the reason for distinguishing these two questions is not that their answers are going to diverge wildly.[21] Indeed, it would be strange, if not incoherent, if they did: if the answer to 2 is "yes," then it follows as a matter of logical entailment that, in general, the corresponding answer to 1 is "yes."

The reason for distinguishing these two questions is that their answers can diverge—and, more, they almost certainly will diverge in some cases. The answer to 1 sometimes will be "no" even if the answer to 2 is "yes," and vice versa. This is important because Thomson, and all of the Kantian trolleyology that has been written in response to her, seems to presuppose that these two questions do not come apart or that they come apart only in very specific ways, using a model that, I want to argue now, is at odds with Kantian ethics. To see this, we must return to Kleingeld.

Recall that, on Kleingeld's account, in Loop, saving the five can be performed on the basis of a maxim that violates the Kantian prohibition, but it also can be performed on the basis of a maxim that is in accordance with the Kantian prohibition. In explaining Kleingeld's solution in section 3 above, I left things there. However, the crucial step in Kleingeld's solution actually comes next. Kleingeld distinguishes between actions and maxims, and she argues that an action is permissible if, but only if, it can be performed on the basis of a permissible maxim. From this, according to Kleingeld, it follows that the *action* of saving the five in Loop is permissible, as per Thomson's (original) intuitions; we just have to bear in mind that a permissible action "remains permissible even when it is performed on the basis of a morally impermissible action principle" (Kleingeld, 2020b, p. 223). Kleingeld uses this to explain how Loop differs from Fat Man:

> [T]he action of pushing the heavy man off the bridge (given the details of the case) *necessarily* involves the agent's using him merely as a means...For an action to be morally permissible, it should, of course, at least be possible for an agent to perform it without violating moral constraints. Thus, if an action can be performed only on the basis of morally impermissible action principles, then it is impermissible (Kleingeld, 2020b, p. 225).

That is, the action of saving the five is permissible in Loop, for, in that case, it can be performed on the basis of a permissible maxim, but it is impermissible in Fat Man, for, in that case, it can be performed only on the basis of impermissible maxims.

I would like to say four things about this.

[21] A third question that needs to be asked, one that Wood does a nice job of discussing, is: What should someone say when asked about 1 or 2 (Wood, 2011).

First, this account of the action-maxim distinction, and how this distinction can be used to make sense of intuitions about the im/permissibility of actions, is prominent elsewhere in discussions of Kant's and Kantian ethics. For example, in response to objections that Parfit raises in *On What Matters*, both Pogge and Nyholm independently advance the same account as Kleingeld's: they argue that Parfit's intuitions, which Parfit presents as in conflict with Kant's ethics, may be seen to be in conformity with it once we understand the action-maxim distinction and, more specifically, that an action is impermissible if, but only if, it can be performed only on the basis of impermissible maxims (Nyholm, 2015; Parfit, 2011; Pogge, 2004).[22] The model with which Kleingeld is working is widespread in Kantian ethics.

Second, when Kantians champion this model, they often do so on the basis of Kant's remarks, in part I of the *Groundwork to a Metaphysics of Morals*, about action in conformity with but not from duty. For example, consider the following excerpt from Kleingeld:

> It is a familiar idea within Kantian ethics that a permissible action can be performed on the basis of permissible or impermissible action principles ('maxims'). A helpful illustration of this point is Kant's well-known *Groundwork* example of a shopkeeper who charges children the right price on the basis of an impermissible action principle, namely on the basis of the maxim of only pursuing his own long-term interests (G 4: 397). Charging children the right price (the action), Kant says here, is 'in accord with duty,' regardless of whether the shopkeeper's maxim satisfies moral requirements (Kleingeld, 2020b, p. 223).

On Kleingeld's reading of Kant, a shopkeeper who gives correct change from prudence rather than from duty is performing a permissible action from an impermissible action principle.

However, this reading is strained. On the one hand, Kant thinks that giving correct change is obligatory—this follows from the fact that the action can be performed from duty. Of course, there is every reason to think that, on Kant's account, an obligatory action is, *a fortiori*, permissible. But, an obligatory action is not merely permissible, and that is the deontic status Kleingeld seems to want to assign to saving the five in Loop. On the other hand, there is also every reason to think that, on Kant's account, the shopkeeper's maxim (and, *eo ipso facto*, the shopkeeper's act token—more on this in a moment) is permissible. Kant thinks that action from self-interest is permissible so long as this does not conflict with morality[23]. Of course, if the shopkeeper is unwilling to give the correct change *except* when it is prudent to do so, this would exhibit a fault of character on Kant's account. But, we are not given

[22] This is discussed at greater length in section 2 of Kahn (2021).

[23] For helpful discussion, see Wood (1999, chapter 1, esp. section 3.2).

enough information to make any pronouncements about that, nor is it relevant to Kant's discussion at this junction. In sum: this example does not support the model that Kleingeld and others pull from it.

Third, there is independent textual evidence against ascribing this model to Kant. For example, in the *Groundwork to a Metaphysics of Morals*, Kant asks whether it is permissible for one imagined agent to commit suicide in trying circumstances and whether it is permissible for another imagined agent to tell a lying promise in a case of financial distress, actions that Kant evidently thinks are generally impermissible. In both cases, Kant's agents come to the conclusion that the maxims on which they propose to act are contrary to duty. But, the question would not even arise if these impermissible actions could not be performed on the basis of permissible maxims, or if the only way to know whether an action is impermissible is via the knowledge that every maxim on which it can be performed is impermissible. Similarly, in the *Metaphysics of Morals*, after explaining why various actions, like suicide, are impermissible, Kant poses casuistical questions, which suggests that, on his account, the deontic status of an action type does not logically entail the deontic status of tokens of the type: permissible action types can be performed on the basis of impermissible maxims, as Kleingeld asserts, but, in addition, impermissible action types can be performed on the basis of permissible maxims (*pace* Kleingeld et al).

Fourth, this model is philosophically problematic. For one thing, because any action type can be instantiated in infinitely many tokens, and because there does not seem to be any way to assess these infinitely many tokens except by examining them one by one, given our limited capacities, cognitive and otherwise, nobody ever will be in a position justifiably to assert that a given action type is impermissible[24]. In other words, this model makes impermissibility into an unusable category. For another thing, because so many intuitively impermissible action types can be performed on the basis of permissible maxims, this model is going to have many false positives. To take just two examples: if, as many assert, it is permissible to kill in self-defense or to

[24] There is room for some pushback here. If, as Kleingeld has pointed out to me, we individuate action types by the maxims on which they are performed, then we can determine that an action type is impermissible by testing one maxim.

However, there are two problems with this proposal. One is that it presupposes that maxim types, rather than maxim tokens, are the locus of assessment in Kant's ethics, and this presupposition has recently been called into question. The other, which is independent of the first, is that action types are generally not individuated in this way (e.g., whether Emily lies in order to get revenge, or whether she lies in order to get some ready money, she is telling a lie—the action kind (lying) remains constant through alternate maxims), and for a good reason: because action kinds, unlike action tokens, can be performed on infinitely many maxims, this would render the moral landscape too fine-grained for pronouncements about action kinds to be practicable (indeed, it would fail to advance us in the trolley problem, precisely because the action kinds in question—pulling the switch, shoving the fat man, etc.—can be performed on a multitude, if not an infinitude, of maxims, something that Kleingeld herself presupposes in her solution to the trolley problem when she argues that pulling the switch in Loop can be performed on two alternate maxims).

lie to a murderer from philanthropic motives, then killing and lying are permissible action types on this model—and this seems like a deeper problem than the trolley problem which this model is supposed to solve[25]. This model arguably arises from a confusion of action types and action tokens, for, on Kant's account, an act token is impermissible if, but only if, it is performed on the basis of a maxim that violates the Categorical Imperative. So, let me conclude this article with a brief summary of the model I advocate and the direction it points us in regard to the trolley problem.

On my reading of Kant, he has a tripartite distinction: action tokens, action types, and maxims[26]. Although Kant does not address the question in these terms, I believe that he would say that an action token is metaphysically individuated, at least in part if not in whole, by the maxim(s) on which it is performed. This is important because it cuts off at its root any attempt to make the deontic status of an action token a modal property based on whether it *could* be performed on a maxim with a deontic status that differs from that of the maxim on which it actually is performed. If I am right about this, then no act token could be performed on any maxim other than the one(s) on which it actually is performed, whence it follows that no act token could be performed on any maxim with a different deontic status.

Support for this derives from Kant's formulations of the Categorical Imperative. For example, consider the Formula of Universal Law as articulated in part II of the *Groundwork to a Metaphysics of Morals*: "Act only according to that maxim through which you can at the same time will that it become a universal law" (GMS, AA 04: 421.07-08, emphasis omitted). From this formulation, it may be seen that, on Kant's account, the permissibility of an act token is to be determined by appeal to the maxim on which it is performed—not by appeal to a possible maxim on which it can be performed. Maxims are either universalizable or not, or they involve using humanity at the same time as an end or not, and act tokens are im/permissible based on this.

Turning to act types: as I read Kant, act types form the content of general duties, as argued for and categorized in the *Metaphysics of Morals*. To determine the deontic status of an act type, we can make generalizations about humans and our circumstances in order to arrive at plausible generalizations about the kinds of maxims on the basis of which we perform (or omit) act tokens of the act type in question[27]. We then say that the type is permissible if, but only if, tokens of the type are generally permissible (i.e., if, but only if, tokens of the type are generally

[25] Indeed, the model will undermine Kleingeld's solution to the trolley problem inasmuch as it undermines her pronouncement about the Fat Man case.

[26] This distinction has recently been emphasized in Kahn (2024b).

[27] For ease of exposition, I proceed without reference to duties of omission.

performed on the basis of universalizable maxims)[28]. This puts me at odds not only with the Kantian approaches to trolleyology canvassed in the foregoing, but also, more broadly, with many Kantian approaches to general duties. For example, on my account, general duties do not form deliberative presumptions that must be rebutted (*pace* Herman, 1993, chapter 7), nor are they provisionally universal, with exception clauses to be worked into them as we encounter them (*pace* Korsgaard, 2002, section 2.5.2; see also Korsgaard, 2008, p. 122)[29]. Rather, general duties are approximations, the ethical equivalent of back-of-the-envelope calculations, and this means, again, not only that a permissible act type can be tokened impermissibly (as Kleingeld, Nyholm, and Pogge assert), but also that an impermissible act type can be tokened permissibly (*pace* Kleingeld, Nyholm, and Pogge). Moreover, I think that textual evidence for ascribing this view to Kant can be derived, not only from Kant's use of casuistical questions, as described above, but also from Kant's own description, in the *Metaphysics of Morals*, of how general duties are to be argued for:

> [W]e will often have to take the special nature of humans, which is cognized only through experience, as an object in order to show in it the consequences from universal moral principles, without that, however, thereby taking away anything from the purity of the latter, or its a priori origin being made thereby doubtful. —That is as much as to say: a metaphysics of morals cannot be grounded on anthropology, but nevertheless [it can] be applied to it (MM, AA 06: 217.01-08).

According to Kant, in order to derive general duties, we have to make generalizations about humans and, more specifically, about the kinds of principles we espouse in tokening certain types, in order to determine the deontic status of those types. So, how is this model to be applied to the trolley problem?

In order to determine the deontic status of saving the five in Trolley, Fat Man, or Loop, we need to make plausible generalizations about the kinds of maxims on which agents in these situations would perform these act types. But, there is a subtlety that is worth remarking upon: the idea is not to articulate the maxims that (we stipulate) are *possible*, as per Kleingeld, but, rather, to articulate the maxims that agents generally actually would adopt. The first complication that arises, then, returns us to the Wood/Fried critique: it is unclear whether most agents, thrust into such a situation, would act unreflectively, on the basis of previously adopted maxims, or whether they would have the time and wherewithal to adopt a maxim tailored to

[28] Other deontic categories, like the obligatory, are arrived at in a slightly different, and sometimes somewhat more circuitous, fashion. Again, for ease of exposition, I omit these details.

[29] General duties might be said to establish an evaluative presumption, where an agent who performs a token of some type is presumed to have behaved im/permissibly until proven otherwise, but that is not quite the same as a deliberative presumption, where an action is presumed im/permissible in an agent's deliberations until she sees her way to performing it from the appropriate motive to make it otherwise.

the situation. To see how these come apart, note that many agents are habituated, and actively habituate themselves, to avoid physical contact with strangers, especially when this involves doing the latter harm: for many, the thought of shoving a fat man off a bridge into the path of a runaway trolley will not even be on the radar. Curiously, this means that the maxims on the basis of which we assess these actions might have little to do with the specifics of the situations at hand.

A second complication that arises is that it is unclear whether agents would know how to divert or stop the trolley. In Trolley and Loop, it is unclear how agents would figure out how to get the trolley onto the other track, even if they were able to see that there are five stuck on the one and only one on the other. In Fat Man and Loop, it is unclear how agents would figure out that another person could be used to stop the runaway trolley. Indeed, it seems *prima facie* implausible that a runaway trolley could be stopped as proposed in the thought experiments—it seems much more likely that this merely would create another victim.

Now, trolleyologists might point out that these complications have been stipulated away: they stipulate the details of the case, right down to what agents do and do not know.

In response, we can echo, once again, the complaint made by Wood and Fried that these stipulations are unrealistic. In order to figure out what the Kantian prohibition would say about these cases, we need to make generalizations about what maxims agents in these (distant) possible worlds would adopt and, then, about whether these maxims would involve using anybody as a mere means. But, because these worlds are so distant, it is exceedingly difficult to see how to do so, and this makes it even more difficult to see how any intuitions we might have about these cases can be brought to bear on principles.

If we want to engage in casuistry in order to sharpen our understanding of general principles of duty (both of what they require and of when they apply), we cannot help but confront difficult cases[30]. But, if the arguments in this article withstand critical scrutiny, then it is at least unclear whether using trolley problems is apt for this purpose.

ACKNOWLEDGMENTS

I am profoundly grateful to the discussants at, and especially the organizers of, the September 2024 conference on "Kant on Means, Ends, and Trolleys."

[30] I would like to thank an anonymous reviewer for pointing this out (I have borrowed some of the reviewer's prose in making the point here).

REFERENCES

Fried, B. (2012). What *Does* Matter? *The Philosophical Quarterly, 62*(248), 505–529.

Friedman, A. (2002). *Minimizing Harm*. Retrieved from: https://dspace.mit.edu/handle/1721.1/8155

Guyer, P. (2006). *Kant*. New York: Routledge.

Herman, B. (1993). *The Practice of Moral Judgment*. Cambridge: Harvard University Press.

Kahn, S. (2013). Review of Kerstein, *How to Treat Persons*. *Kantian Review, 19*(2), 319–323.

Kahn, S. (2021). Obligatory Actions, Obligatory Maxims. *Kantian Review*, *26*(1), 1–25. https://doi:10.1017/S136941542000028X

Kahn, S. (2023). Kant and the Trolley. *Journal of Value Inquiry, 57*, 487–497. https://doi.org/10.1007/s10790-021-09838-6

Kahn, S. (2024a). Consent and the Mere Means Principle. *Journal of Value Inquiry, 58*, 515–533. https://doi.org/10.1007/s10790-022-09909-2

Kahn, S. (2024b). Individual Maxim Tokens, not Abstract Maxim Types. *Kantian Review*. Published online: 1–17. https://doi.org/10.1017/S1369415424000219

Kerstein, S. (2013). *How to Treat Persons*. Oxford: Oxford University Press.

Kleingeld, P. (2020a). How to Use Someone 'Merely as a Means'. *Kantian Review*, *25*(3), 389–414.

Kleingeld, P. (2020b). A Kantian Solution to the Trolley Problem. In M. Timmons (ed.), *Oxford Studies in Normative Ethics, 10* (pp. 204–228), Oxford: Oxford University Press. Accessed from philpapers: https://philarchive.org/ archive/KLEAKS

Korsgaard, C. (2002). *Self Constitution*. Locke Lectures at Oxford.

Korsgaard, C. (2008). *The Constitution of Agency*. Cambridge: Cambridge University Press.

Mahon, J. (2021). Murderer at the Switch. In C. Tandy (ed.) *Death and Anti-Death, Vol. 19* (pp. 153–187), Michigan: Ria University Press.

Nyholm, S. (2015). Kant's Formula of Universal Law Revisited. *Metaphilosophy*, *46*(2), 280–299.

Parfit, D. (2011). *On What Matters Volume 1*. Oxford: Oxford University Press.

Pogge, T. (2004). Parfit On What's Wrong. *The Harvard Review of Philosophy*, *XII*(1), 52–59.

Schmidt, E. (2022). Kant on Trolleys and Autonomous Driving. In H. Kim & D. Schönecker (Eds.), *Kant and Artificial Intelligence* (pp. 189–222). Berlin: De Gruyter.

Seymour Fahmy, M. (2021). Shadow students in Georgia. *Journal of Philosophy of Education, 55*(6), 1057–1071. https://doi.org/10.1111/1467-9752.12609

Seymour Fahmy, M. (2023). Never Merely as a Means. *Kantian Review*, *28*, 41–62.

Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, *59*(2), 204–217.

Thomson, J. J. (1985). The Trolley. *The Yale Law Journal*, *94*(6), 1395–1415.

Thomson, J. J. (2008). Turning the Trolley. *Philosophy & Public Affairs*, *36*(4), 359–374.

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

Williams, B. (1970). The Self and the Future. *The Philosophical Review*, *79*(2), 161–180.

Wood, A. (1999). *Kant's Ethical Thought.* Cambridge: Cambridge University Press.

Wood, A. (2008). *Kantian Ethics*. Cambridge: Cambridge University Press.

Wood, A. (2011). Humanity as an End in Itself. In D. Parfit, *On What Matters, Vol. II* (pp. 58–82). Oxford: Oxford University Press.

Kantovska trolejologija (engl. *trolleyology*)

SAŽETAK

Rad istražuje šest suvremenih pristupa poznatom „problemu trolejbusa“ (engl. *trolley problem*) Judith Jarvis Thomson, misaonom eksperimentu koji i dalje predstavlja izazov i filozofima i nefilozofima. Analizirani pristupi uključuju one Paula Guyera, Pauline Kleingeld, Samuela Kersteina, Elke Elisabeth Schmidt, Jamesa Edwina Mahona i Allena Wooda. Svi ovi filozofi pokušavaju pomiriti Kantov moralni okvir s jedne strane s etičkim problemima koji su proizašli zbog problema trolejbusa s druge strane. Rad, međutim, ističe i objašnjava značajne izazove u pristupu svakog pojedinog filozofa, ukazujući na moguće nedostatke i nedosljednosti. U kritičkom bavljenju ovim raznolikim kantovskim pristupima rad ne samo da naglašava njihova ograničenja, nego i ponovno ispituje Thomsonov izvorni okvir problema kako bismo procijenili trajni značaj potonjeg i uronili u neka od sistemskih pitanja svojstvenih trolejologiji. Tvrdi se da je prepoznavanje ovih sistemskih problema bitno za unaprjeđenje robusnijeg kantovskog rješenja za problem trolejbusa, rješenja koje ostaje vjerno kantovskim načelima, dok se usprkos tome bavi složenošću moralnog odlučivanja u takvim situacijama.

Ključne riječi: trolejologija, kantovska trolejologija, Kantova etika, kantovska etika, Thomson, dužnosti.