

## TROLLEY CASES AND BEING “IN THE REALM”

Michael Barnwell

In “Killing, Letting Die and the Trolley Problem,”<sup>1</sup> Judith Thomson discusses what I will call “two-option threat distribution” cases (TOTDs): cases in which a non-threatened agent must act or not so as to distribute a threat in one of two ways.<sup>2</sup> After observing that in some TOTDs we intuitively approve of minimizing harm while in others we do not, Thomson offers a distinction to explain our divergent judgments. She notes that “what matters in these cases in which a threat is to be distributed is whether the agent distributes it by doing something to it, or whether he distributes it by doing something to a person.”<sup>3</sup>

In this paper I have a two-fold aim. First, I will demonstrate that this distinction is not necessarily “what matters” in TOTDs since it fails to explain our intuitions in TOTDs she does not consider. I will specify cases in which we intuitively opt not to minimize harm even though, if Thomson's suggestion holds, we should. Second, I will argue that another distinction, that between being “in” and “out” of “the realm” may actually be “what matters” in TOTDs.

Thomson develops her distinction by referring to the following two now-infamous thought experiments: SURGEON and RUNAWAY TROLLEY.

**SURGEON:** “David is a great transplant surgeon. Five of his patients need new” organs. He has a healthy patient whose organs he can remove in surgery and transplant into the five sick patients, thereby saving their lives but killing the healthy patient. David operates on his healthy patient and kills him.”<sup>4</sup>

**TROLLEY:** Passenger Frank can control a runaway trolley. Five people are on the track ahead and will be run over by the trolley. Luckily, “the track has a spur leading off to the right, and Frank can turn the trolley onto it. Unfortunately there is one person on the right-hand track.” Frank turns the trolley to the right.”<sup>5</sup>

In both cases, the agent causes one death so as to save five lives. Under the assumption that all six potential victims are equivalent in all relevant aspects, one may initially suspect that either both acted morally or both acted immorally. Nonetheless, it is generally held (by deontologists) that David acted wrongly while Frank did not.

Our intuitions are similarly paradoxical in the following two cases: D-WORCESTER and B-WORCESTER:

**D-WORCESTER:** President Harry “has just been told that the Russians have launched an atom bomb towards New York. The only way in which the bomb can be prevented from reaching New York is by deflecting it . . . onto Worcester.” Harry deflects the bomb.”<sup>6</sup>

**B-WORCESTER:** Similar to D-Worcester, except Irving is President and deflection is not possible. The only way to save New York is to drop “one of our own atom bombs on Worcester,” the blast from which will “pulverize the Russian bomb.” Irving bombs Worcester.”<sup>7</sup>

Typically, Harry's action is regarded as acceptable while Irving's is not.

As noted above, Thomson suggests that our paradoxical judgments in these cases can be explained by distinguishing between “whether the agent distributes [a threat] by doing something to it, or whether he distributes it by doing something to a person.”<sup>8</sup> In TROLLEY and D-WORCESTER, minimizing harm requires only doing something to the threat by diverting it in some manner. In SURGEON and B-WORCESTER, by contrast, minimizing harm requires that the agent do something more directly to people (operate on him, bomb them) and is for that reason proscribed.

As Thomson admits, this distinction is a bit unclear.<sup>9</sup> After all, it seems that in TROLLEY Frank does do something to the one by turning the trolley on a track that ultimately leads to her. Nonetheless, the distinction seems to rest upon a difference between diverting an already-existing threat and preventing an already-existing threat by bringing a new threat to bear upon persons: Frank and Harry can act since they need only divert already-existing threats, whereas David and Irving should not because, to do so, they would need to bring to bear upon victims a new threat (surgery, bombing).<sup>10</sup>

My claim is that this distinction is inadequate to explain our intuitions in all TOTDs and, as such, may not be that which underlies our judgments. Consider:

**BRIDGE-PIECE:** A runaway trolley is racing down the tracks toward five innocents. Construction of an automobile bridge over the tracks is occurring between the trolley’s location and the five. The supports for the bridge are in place and a bridge-piece, currently suspended by a crane, is about to be lowered onto them. A workman is on that bridge-piece. It is possible for the crane operator, Jenny, to lower the bridge-piece not onto the supports, but instead onto the tracks themselves so as to cause the trolley to derail and slide down the bridge-piece away from the five. The workman on the bridge-piece, however, would be killed by the trolley sliding down it.<sup>11</sup>

BRIDGE-PIECE is largely similar to TROLLEY in that, if Jenny were to act, she would be diverting the threat. Moreover, she does not seem to be “doing anything” to the workman on the bridge-piece in that she is not initiating a new threat that is then brought to bear upon the workman. By Thomson’s distinction, therefore, it appears that Jenny should lower the bridge-piece onto the tracks. I doubt, however, that this is the common intuition.<sup>12</sup>

Further criticism of Thomson’s distinction comes from:

**YARD:** An empty trolley is racing down a track toward five innocents. Their only hope is if Abigail flips a switch which will make another empty trolley collide with the first one, causing both to derail and saving the five. However, the derailed trolleys will “go down a hill, across a road, and into someone’s yard, where they’ll wreak havoc on the yard’s owner, asleep in his hammock, as well as many of his bushes.” Abigail flips the switch.<sup>13</sup>

By diverting the threat, it is difficult to see how Abigail could be viewed as doing something “to a person” in Thomson’s sense. She certainly no more does something “to a person” than Frank does in TROLLEY; she does not initiate a new threat against the sleeper in the hammock in the way that David does in SURGEON. It therefore seems that Thomson’s theory should permit Abigail to so minimize harm. As Unger points out,

however, this is not the common intuition: "here most respond that [Abigail's] conduct was wrong."<sup>14</sup> YARD thus casts doubt upon Thomson's distinction.

In order to make my own positive suggestion that the critical distinction in TOTDs is that between being what I call "in" and "out" of "the realm," consider LION and LION2.<sup>15</sup>

**LION:** An evil sorcerer has caged 196 innocents in two cages (A and B) on a beach. One hundred are in A, and ninety-six in B. The cages are in sight of each other. Each cage has a small doorway through which none can escape but a lion can enter. A lion is placed on the beach between the two cages in such a way that both cages are out of its direct sight. To see either cage, the lion must turn his head to the left (to see A) or to the right (to see B). Furthermore, this lion is bewitched such that, whenever caged humans are in its sight, it will attack and kill all those in that cage. The lion, though, will not attack two cages; after having attacked and killed those in one, it will never attack another. The sorcerer has also arranged it such that the lion will turn its head left or right and see the corresponding cage. It follows that the lion will then keep that cage in sight and kill those in that cage. It is not yet determined whether the lion will turn left or right. The sorcerer has, however, made it such that the lion will see and kill all of the persons in one of the cages, and the kidnapped innocents know this.

The lion turns left and sees A which contains one hundred persons. Accordingly, the lion proceeds toward A in order to enter it and kill its one hundred inhabitants, causing great relief for the ninety-six trapped in B who watch this. Alex, also knowing the details of the situation, somehow arrives on the scene just as the lion has gone through the small doorway of A but has not yet killed anybody. Alex cannot reverse the spell that dictates all the inhabitants of one of the cages will be killed. However, he can somehow reach through the small doorway, pull the lion out, and (knowing that at least one of the cages must be attacked) turn its head so that it sees B and attacks its ninety-six (instead of one hundred) inhabitants. Alex decides to minimize harm by doing so.

**LION2:** Same as LION, except Steve arrives before the lion has initially turned its head. Knowing that one of the cages must be attacked, Steve decides to turn the lion's head toward Cage B so that four fewer people die.

It is not necessarily clear, to me at least, whether Steve acted correctly; it may be that he was morally required to do nothing. What is clear, however, is that Steve's action is not as objectionable as Alex's. Given that both Alex and Steve minimized harm by turning the lion onto B, what could account for the different intuitive reactions to their acts?

I propose that our intuitive response to Alex's action is grounded upon his violating a constraint against putting somebody "in the realm" of a threat if they are "out of the realm." To explain, I must first discuss the further concept of a "normal tipping point" (NTP). By NTP, I understand a point at which there are two or more viable alternatives with respect to a threat and at which, given the way things "normally" are, the threat could *just as likely* be "tipped" to one option as to the other without any undue influence.

Given this understanding of NTP, one is “in the realm” if one is potentially subject to a threat and either (i) there remains a NTP with regard to the threat and the threat could be “tipped” toward her at that NTP, or (ii) the threat has already been “tipped” toward her either at a NTP or an ATP (“abnormal tipping point”). By contrast, one is “out of the realm” not only if she is not potentially subject to a threat, but also if she is potentially subject but the threat has been “tipped” away from her and there remain no more NTPs (though there may remain ATPs) at which it could be “tipped” toward her.<sup>16</sup>

An initial understanding of NTPs and “the realm” can be gathered from LION. The lion’s NTP is the point at which it looks, of its own accord, at one cage or another and “zeroes-in” (so to speak) on that cage. In LION, the lion saw and “zeroed-in” on A at its NTP. Since it had already “zeroed-in” on A, it would not of its own accord kill those in B. The lion could have zeroed-in on B at its NTP. But when it “zeroed-in” on A, it thereby “un-zeroed-in” on B. So far as the lion was concerned, those in B were then safe. In other words, there were no more so-called “normal tipping points” with regard to the lion and its actions by which those in B could come into danger. The latter were thus “out of the realm.” When Alex pulled the lion out of the cage and pointed it toward B, he thereby put those in B “in the realm” of the threat after they were out of it and thus broke our proposed constraint. Contrast this with LION2. There, the NTP of the lion zeroing-in on a cage has not passed; the threat could still be tipped toward those in B at a NTP. Accordingly, those in B are not yet “out of the realm.” Steve’s action, therefore, does not violate the same constraint of placing those out of the realm into the realm.<sup>17</sup>

LION and LION2 are, of course, not normal. This, however, has no relevance to the existence of “normal” tipping points in the way here understood. By the latter use of “normal,” I mean to pick out tipping points that are, in some sense, concordant with the nature of the threat itself. Alex’s decision to redirect the lion is, of course, a tipping point in that the threat could then be “tipped” toward different victims. But it does not appear to constitute a “normal” tipping point with respect to the threat itself; it is an *abnormal* tipping point (ATP) given this particular threat. Here the threat is a lion bewitched to look at a cage and attack *that* cage, thereby bypassing another cage. Given this threat, it is not “normal” for it to be redirected by an outsider, not attack the first cage, and attack the second cage, of itself. Alex’s action thus constitutes a sort of undue influence in this situation.

As further clarification, recall SURGEON. There, the healthy patient is clearly out of the realm of the danger of organ failure. The threat of organ failure currently pertaining to others does not “normally” strike one by means of a surgeon removing one’s own healthy organs. There are thus no NTPs by which that threat could be “tipped” toward the patient. To cause organ failure in the healthy patient by surgically removing his organs was to tip the threat toward the patient at an ATP, not a NTP. Consequently, David’s immoral action can be explained by our distinction: he caused the patient to enter the realm of the threat when he was out of it.

In TROLLEY, the racing trolley is the threat. Trolleys, by their nature, are just as apt to take the right as the left fork as they are rolling down the track. Assuming the fork does not lead to a circular track, then once the trolley has taken one fork it has thereby not taken the other fork. A fork in the track, therefore, is a NTP.<sup>18</sup> This helps explain why

Frank can take the right fork. Before the trolley passed its NTP of the fork, the victim on the right was still in the realm and never out of it; Frank did not move a victim from out of the realm into it by taking the right fork.<sup>19</sup>

By contrast, trolleys are not “just as likely as not” to jump tracks onto bridge-pieces. To do so is not “natural” or “normal” for trolleys, even runaway ones. The worker on the bridge-piece, therefore, is out of the realm; there is no NTP at which the threat of the trolley could be tipped toward him. For Jenny to lower the bridge-piece in BRIDGE-PIECE, therefore, would break our constraint of moving somebody who is out of the realm into it.

YARD can be similarly explained by the proposed constraint. Trolleys are not apt to be used as derailment mechanisms for other trolleys. Nor are trolleys “just as likely as not” to be collided into by another trolley and derailed. Of course, such can happen as it does in YARD. But such an occurrence is not a “normal” one, even for a runaway trolley. Consequently, there is no NTP at which the threat of the trolley could be tipped toward the homeowner asleep in his hammock. As such, he is out of the realm. Abigail consequently breaks our constraint by putting him “in the realm” by means of an ATP.

It might be thought that the reasoning offered so far makes it difficult to explain why Harry can deflect the missile to Worcester but Irving cannot bomb Worcester. After all, the missile has already been launched toward New York and has “zeroed-in” on that city, so to speak. In response, note that even though (a) it is *not* “normal” for bewitched lions to change their focus once they have “zeroed-in” on a cage, and (b) trolleys are *not* “just as likely as not” to be collided into and derailed by other trolleys, it (c) *is* “normal” and “just as likely as not” for hostile missiles to be deflected. Indeed, it is in the nature of missile warfare to attempt to shoot down and/or divert attacking missiles. Such points at which the missile could be deflected, therefore, are NTPs. In D-WORCESTER, Worcester is a city toward which the threat of the missile could be “tipped” at one of these NTPs. Thus, it is always in the realm. Accordingly, Harry does not break our constraint by deflecting the missile. In B-WORCESTER, however, there either never were or are no more NTPs with regard to the threat at which it, the missile, could be “tipped” toward Worcester. Accordingly, Worcester is “out of the realm.” For Irving to bomb it, therefore, would be to place Worcester in the realm of danger related to the missile after it was out of it.

We have seen that Thomson’s distinction does not necessarily explain our intuitions in TOTDs. I have accordingly offered another: that between being in and out of the realm. We intuitively disapprove of minimizing harm if it requires placing into the realm of a threat one who is out of it.<sup>20</sup> This distinction, moreover, rests upon an understanding of NTPs. One is out of the realm if there are no more NTPs at which the threat could be tipped toward oneself; one is in the realm if either the threat is already “tipped” toward one, or there is a NTP at which it could be so tipped.<sup>21</sup> While more must be done so as to specify what counts as normal (as opposed to abnormal) tipping points with regard to various given threats,<sup>22</sup> this distinction permits an alternative understanding of the trolley problem and when it may or may not be permissible to minimize harm.

## NOTES

1. Judith Jarvis Thomson, "Killing, Letting Die, and the Trolley Problem." *Rights, Restitution and Risk*. (Cambridge, Mass: Harvard, 1986).

2. Please note that I am here excluding from consideration ducking vs. enshielding cases such as those discussed in Christopher Boorse and Roy A. Sorenson, "Ducking Harm," *The Journal of Philosophy*, 85 (1988): 115-134. In the ducking/enshielding cases, there is not an agent who is not a potential victim distributing a threat as there is in TOTDs. Furthermore, please note that, by focusing on TOTDs, Unger's multi-option threat distribution cases will not fall under consideration in this paper: see Peter Unger, *Living High and Letting Die* (New York: Oxford UP, 1996).

3. Thomson 92.

4. Thomson 80. I have slightly emended the thought experiments Thomson presents.

5. Thomson 81-82.

6. Thomson 83.

7. Thomson 83.

8. Thomson 92.

9. Thomson, 92.

10. For a similar parsing yet different discussion of Thomson's suggested distinction, see John Martin Fisher and Mark Ravizza, "Thomson and the Trolley," *Journal of Social Philosophy*, 23.3 (1992): 115-134.

11. The basic outlines of this variation, some of which I have slightly altered, were suggested to me by Shelly Kagan.

12. And even if a few do believe Jenny can so act, such are no doubt fewer than those who believe Frank can act in TROLLEY. I further suspect that any who believe Jenny can act do not believe it to the degree they believe Frank can, which they should if Thomson's distinction were the relevant one underlying our intuition.

13. Unger 98, proposes this thought experiment which I have slightly modified here. So far as I can tell, Unger has a different purpose for this thought experiment than we do; he is not necessarily addressing Thomson's distinction.

14. Unger 98.

15. LION was inspired by examples in Boorse and Sorensen's "Ducking Harm." It is here, however, drastically different from, and used for a different purpose than, any of their examples.

16. Note that if one is out of the realm yet still potentially subject to the threat, then there must remain an ATP at which that threat could be tipped toward her. All acts of placing one who is out of the realm into it must, by definition, take place at ATPs.

17. Again, it is acceptable to remain agnostic as to whether Steve's action was itself moral or not. It is possible that he should have left the lion alone. If Steve acted immorally, though, it is not simply because he

## Trolley Cases and Being “in the Realm”

violated the same constraint that Alex did. Moreover, if Steve did violate a constraint, it must be one Frank somehow does not violate in TROLLEY under the assumption that Frank’s turning the trolley is permissible. The only point being made here is that we judge Steve’s and Alex’s actions differently, and (I claim) we do so because Alex breaks a constraint that Steve does not, namely the constraint of placing ones who are outside the realm into it.

18. Note that trolleys are not apt to be lifted off of the right track to the left track. Trolleys are made to roll on the tracks and take the forks by rolling. Consequently, once a fork has been taken by a trolley rolling onto it, the trolley has passed its NTP. If there were a point after the fork at which a trolley were lifted and placed onto another track, such an abnormal tipping of the threat to those on the other track would constitute an ATP, not a NTP.

19. Although my conclusion and purpose are different, the present understanding of being in and out of the realm might help elucidate C.L. Sheng’s discussion of what he calls a “pre-action situation” and his contrast between a “normally expected” and “another kind of contingency.” See C.L. Sheng, “A Suggested Solution to the Trolley Problem,” *Journal of Social Philosophy* 25.1 (1995): 203-217 (esp. 210).

20. It is interesting to note that, at times, Thomson attempts to parse out the relevant distinction in TOTDs in terms of whether persons have “claims” against a threat. Being “out of the realm” could be viewed as one’s having a “claim” against a threat. Consequently, the current proposal may be regarded as an attempt to support Thomson’s reference to claims.

21. Note that if only ATPs remain, then one is “out of the realm” as it is understood in this paper even though one is still potentially subject to a threat.

22. What is “normal” would, of course change for each given threat. This, however, is no objection to the proposal offered here. If my proposal were accepted, then moral reasoning regarding TOTDs could be more suitably focused: it would be focused not upon what distinction is underlying our intuitive appraisals, but rather upon the further question of what should count as a “normal” point at which threats could be tipped. As such, the acceptance of the distinction being offered would advance our moral reasoning.