

Towards a Modification of the No-New-Threat Principle

By Matthew Matasar

SINCE THE INTRODUCTION OF UTILITARIANISM AS A philosophy, much of the philosophical community has been split along clearly defined lines—either in favor of or against the philosophical basis and normative conclusions of Utilitarianism. Most recently, the efforts of the anti-Utilitarians have in part been focused on defeating Utilitarianism on the grounds that it fails to capture our intuitions about morality. Much of this effort has been devoted to explaining and supporting the concept of agent relativity. Whereas normative Utilitarianism dictates that those actions which produce the greatest increase in happiness must be pursued, the agent-relativists have responded that it appears to everyone on a basic, intuitive level that “special” cases exist in which it seems at very least morally permissible to go against such utilitarian calculations.

In such cases, the moral status of an action is dependent not only upon the results of the action, but also upon the nature of the action itself—in other words, the moral worth of an action is in part relative to the agent and the action. Beyond this basic phrasing of the principle of agent relativity, however, there is further disagreement, as numerous agent-relative principles have been proposed which are claimed to fully capture our moral intuitions. These principles run the range from conservative to extremist—some have asserted that it is better to allow 5 to die than to actively kill 1, while others have maintained that it is perfectly permissible to choose the life of one friend over the deaths of countless strangers. Two agent-relative principles which lie somewhere between the extremes of the spectrum are of particular interest to me, however, as they seem to capture many of our moral intuitions—the Doctrine of Double Effect and the No New Threat principle.

The Doctrine of Double Effect is not, in fact, a philosophical innovation devised in response to Utilitarianism; rather, it is an old doctrine of the Catholic church which has been revived because it does seem to capture many of our moral judgments. The Doctrine of Double Effect (DDE) states that one may not aim at the death or harm of an individual as an end in itself or as chosen means to another end. While somewhat archaic, when taken as a moral doctrine rather than a religious one, it can explain the permissibility of strange cases which Utilitarianism incorrectly predicts our intuitions on permissibility; this advantage of the DDE stems from its consideration of intention—not only do the results and the means of the action figure into moral judgments, but the reasons behind the action also shape our opinions about permissibility from a moral standpoint.

To more clearly understand the appeal of the DDE as compared to Utilitarianism, let us take two examples offered originally by Philippa Foot, known as the Trolley and the Surgeon scenarios. In the Trolley, you are the

MATTHEW MAT-
*Asar is a freshman
 at Harvard College. He
 intends to concentrate in
 History and Science. He
 would like to thank Leif
 Wenar for his patience.*

“You are an incredibly talented surgeon with five patients who are going to die within the day. Miraculously, a young man who is a perfect tissue donor for all five has just entered your office for a physical...”

driver of a trolley heading downhill, and before you there are five workers on the track. To your dismay, you soon find that the brakes on the trolley are out. However, the track diverges just before the workers, and on the second track, onto which you can turn, there is one worker. It is clear to everyone on an intuitive level that it is at least permissible to turn onto the second track—Utilitarianism can explain this reaction. In the Surgeon, you are an incredibly talented surgeon with five patients who are going to die within the day—two need a lung, two need a kidney, and one needs a heart. Miraculously, a young man who is a perfect tissue donor for all five has just entered your office for a physical. Again on an intuitive level, nobody could be so cold-hearted as to find it permissible to hack up the youth for spare parts — *despite* the fact that it would increase happiness more than letting him walk out of your office and letting the five die. While our moral intuitions in this case are equally clear as those in the Trolley problem, Utilitarianism struggles in its effort to explain the obvious impermissibility of killing the young man.

UTILITARIANISM MANAGES TO EXPLAIN THE difference between Trolley and Surgeon only with great difficulty, if at all; the DDE explains it effortlessly. In the Trolley, your aim is not to do harm to the one, but rather to avoid harming the five. Furthermore, you are not treating the one's harm as a means to another end, for your action would be the same had he not been standing on the second track. In other words, the success of your avoiding harm to the five in no way hinges on the harm of the one. Not so in Surgeon; while the success of your saving your five patients does not exactly hinge on the death of the youth, it certainly does hinge on harming him by cutting him open and taking his vital organs. Thus, the DDE accurately predicts the difference between two situations which are roughly equivalent according to pure Utilitarianism.

The Doctrine of Double Effect, however, does not in and of itself correctly predict all of our moral intuitions. One particular case to demonstrate this point was offered by Judith Jarvis Thomson, the author of the No New Threat principle. Let us return to the Trolley problem of a moment ago. Instead of having the two tracks diverging, let us imagine it as a looped track, a knob. After establishing this variation diagrammatically, Thomson continues,

“Let us now imagine that the five on the straight track are thin, but thick enough so that although all five will be killed if the trolley goes straight, the bodies of the five will stop it, and it will therefore not reach the one. On the other hand, the one on the right-hand track is fat, so fat that his body will by itself stop the trolley, and the trolley will therefore not reach the five. May the agent turn the trolley?” (Thomson, Judith Jarvis, “The Trolley Problem”, *Yale Law Journal*, volume 94, 1985, p. 1403)

According to the Doctrine of Double Effect, the action is no longer permissible, because you are turning into the fat man no longer so as to avoid hitting the five but now so as to stop the trolley. Seen in a different light, the action in the original problem was turning the trolley, whereas in the looped Trolley problem the action is plowing into the fat man. However, it is at least not immediately apparent why you may be allowed to turn the trolley in the original formation of the scenario, but not in this one. Thus, the DDE seems to fall short of capturing our every intuition. In response to the apparent shortcomings of the DDE, Thomson proceeds to offer her own

agent-relative principle, the No New Threat principle, which states that there exists a “distributive exemption...[which] permits intervention into the world to get an object that already threatens death to those many to instead threaten death to these few, but only by acts that are not themselves gross impingements on the...stringent rights of the few” (YLJ, p. 1412). Before we decide just how well the No New Threat principle (NNT) accords with our moral intuitions, it is worthwhile to clarify just what Thomson means to say with this principle, as her terminology allows for multiple interpretations. The two areas of ambiguity are what is intended to be considered an object which already threatens death and what are to be understood as stringent rights. It is somewhat clear that the pre-existence of the threat is not taken to be read too literally—if a murderer were to kill 5 people to his left with his gun unless you told him not to, in which case he would murder one person to his right with a battleax, you do not create a new threat by commanding him, for while the axe threatened nobody, it is the murderer who is the threat in both cases. There is no such clear-cut interpretation of Thomson’s definition of “stringent rights”, unfortunately—while she makes it clear countless times that there is a stringent right to life, she herself states that she does not wish to offer any specific list of stringent rights, instead appealing to an intuitive understanding of which rights are stringent and which are not. While this does not seem wholly satisfactory, we will grant Thomson her intuitive definition of stringent rights, as an extensive rights analysis falls outside of the scope of this paper, and move on.

UPON INSPECTION, THE NNT HAS EXTRAORDINARY success in predicting our moral intuitions. Recalling the looped Trolley problem which appeared to be the downfall of the DDE, it is clear that the NNT captures this scenario. After all, the threat is an existing one—without its brakes, the trolley will kill somebody. Thomson further contends that the mere turning of the trolley is not in and of itself a violation of the fat man’s rights. As this is not immediately obvious, she suggests that we compare our intuitions in this case versus one in which we push someone in front of the trolley. We surely find such an action impermissible, whereas we allow ourselves to turn the trolley. Why this is, she contends, is that in pushing someone into the track of a trolley we violate a stringent right, whereas we fail to do so merely by turning the trolley.

It is obviously essential that the NNT, as a response to the DDE, offer the “correct” response (that is, the response which corresponds to our moral intuitions) to the scenarios for which the DDE failed to do so. The larger question, however, is whether or not Thomson’s principle captures *all* of our moral intuitions. To begin our answer to this larger question, let us return to our original scenarios, the Trolley and the Surgeon, to see if the NNT captures these as successfully as did its predecessor. Sure enough, it discovers an essential difference between the two scenarios, albeit a different one than the DDE identified. In the Trolley problem, the trolley is an existing threat, and no rights are directly violated—the NNT does not draw a moral distinction between the original Trolley problem and the looped Trolley problem, and this coheres to our moral intuitions. I doubt that there are many individuals (besides pure Kantians, perhaps) who would find it permissible to turn in the first case but not in the second. The NNT successfully predicts our prohibition upon action against the youth as the

“If a murderer were to kill 5 people to his left with his gun unless you told him not to, in which case he would murder one person to his right with a battleax, you do not create a new threat by commanding him...”

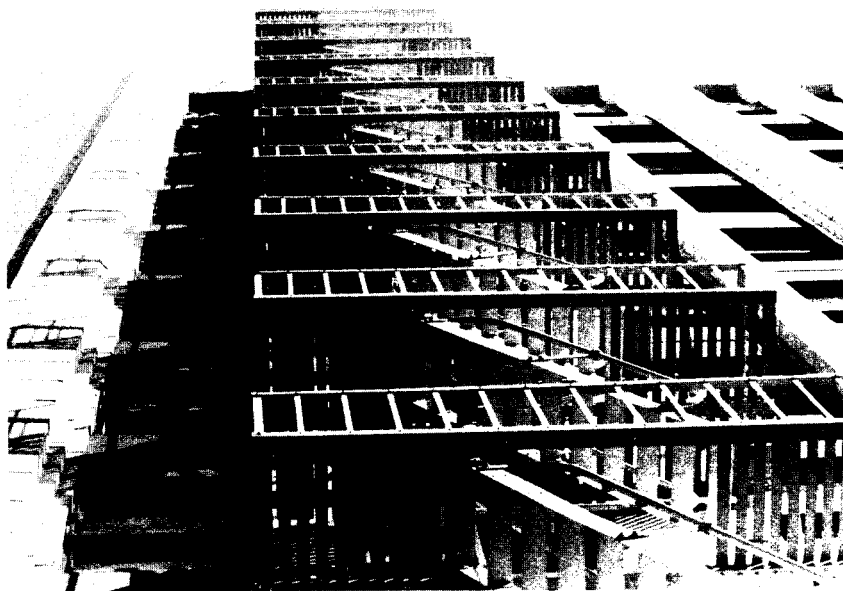
Surgeon, because the existing threat is natural organ failure, as opposed to unsolicited organ removal. Furthermore, it is clear that the youth in your office for a checkup has a stringent right not to have his organs removed against his will; thus, to strip him for his parts is impermissible according to both the NNT and our moral inclinations, and the NNT passes this test.

In fact, there are a vast number of subtle and sticky situations for which the NNT offers answers which coincide with our intuitive judgments. However, a closer and more critical look at the NNT reveals that while it is able to predict responses which the DDE failed to, there are situations in which the DDE successfully predicts our judgments while the NNT fails to do so. For instance, let us consider this example, which I will call “Mine Shaft”.

Imagine that you are a member of a six-individual scientific squad which is to study the air quality in a mine. You and four others have entered the mine, but before your fifth colleague can follow behind you, there is an enormous landslide at the mouth of the mine, trapping the five of you inside the mine—and sealing off the air flow as well. Using your equipment, you discover (to your obvious dismay) that you have a paltry 2 minutes’ worth of oxygen in the mine before the five of you die of asphyxiation. Fortunately, in your emergency supplies is a stick of high explosives, clearly more than enough to blow through the mud and rock blocking the mouth of the mine.

As you position the explosives, you hear your colleague on a megaphone, yelling that there is no need to worry, that he is digging through the material and should have you out in a few hours. You call to him to get out of the way, but he is unable to hear you.

Is it permissible to light the explosives, blowing through the barrier—and killing your colleague in the process? Everyone to whom I have posed this scenario has responded affirmatively—that yes, it is permissible to do so. I will not address the question of whether or not such a course of action is imperative, following Thomson’s precedent. However, it is clear that our intuitions in this case are that it is permissible to kill your colleague. Does the NNT find the scenario permissible as well? I believe the answer to be a resounding “No”. In lighting the stick of dynamite, in fact, you are violating both of the clauses of the NNT: First, you are beyond doubt creating a new threat (the dynamite previously threatened nobody; the source of threat was asphyxiation). Secondly, I believe that a strong case can be made that lighting a stick of explosives close to another person is a violation of his stringent right to life: I leave this open to discussion by rights theorists. Nonetheless, lighting the explosives clearly violates the “no new threat” clause, regardless of the rights issue, and thus the NNT finds this action impermissible. However, the DDE *does* succeed in capturing our judgments in this case. Clearly, the aim of your lighting the dynamite is to free yourselves from the mine, not the cold-blooded murder of your colleague. Thus, as your intention is not blameworthy, the action is permissible according to the DDE.



What do the results of this scenario tell us about the two agent-relative principles? Clearly, they each capture our intuitions in situations in which the other fails to do so. We can, however, arrive at a rephrasing of the NNT to account for the flaws presented by the Mine Shaft scenario. To do so, we must decide what about the scenario is not contained by the NNT. This question, fortunately, has a readily apparent answer—the NNT fails to recognize any influence intention plays in our moral judgments. From this fact, we can arrive at NNT', which would posit:

One may not create a new threat that is intended to inflict harm upon the few as an end in itself or as a means to saving the many; nor may one divert an existing threat from the many to the few in such a way as to grossly impinge upon the stringent rights of the few.

This new expression of NNT, which is in fact a synthesis of the NNT and the DDE, seems to effectively solve the problems which the two agent-relative principles experience independent of one another. In order to make sure this is so, we should decide whether or not NNT' concurs with our intuitions on the looped Trolley problem and the mine shaft problem. With the looped trolley scenario, NNT' clearly offers the correct answer, for we are not creating a new threat which inflicts deliberate harm on the few to spare the many—we are diverting an existing threat to do so. Furthermore, this diversion is not an infringement on the stringent rights of the few, as we have already seen. NNT' also agrees with our intuitions in the mine shaft problem, because while we are creating a new threat, it is not one which *intentionally* inflicts harm upon the few—this is merely a regrettable side-effect of the action, which is clearly intended to open the mine shaft. So it appears that NNT' is a successful synthesis of its parent principles, possessing the strengths of both and the weaknesses of neither.

BEFORE WE CAN CONCLUDE THAT OUR EXPRESSION NNT' is the ideal agent-relative principle, however, we must step back and consider possible objections to it. One convincing criticism takes its form in this scenario, which I will call "Gunman": Now imagine that you are walking down the street when a man walks up to you. He tells you that he is now going to kill the next five people he meets, and that the only way that you could stop him would be to grab one of his guns and kill him. You recall hearing about this man, who has done this before and, when the people in your position chose to let him live, he did in fact kill innocent people. Is it morally permissible to take one of his guns and kill him, even considering the fact that you yourself are at no risk either way? The unanimous consensus which I have heard is that yes, it is permissible in this case to kill him. Again, the question of obligation is beyond my scope; while NNT and NNT' indicate which actions are and are not permissible, it makes no direct claim that the permissible actions are obligatory. Upon analysis, our revised agent-relative principle NNT' still does not seem to capture this scenario. You are creating a new threat (the gunman is the original threat, you are the new threat), and your action is one of intentional harm—it is not a byproduct of another action, but an act of violence in and of itself. Granted, the cause is noble, but whereas in Mine Shaft the success of your act (indeed, your survival) did not hinge upon the death of your colleague, in Gunman you are in fact aiming at the death of another. Thus, according to NNT', the action would be perceived as immoral, while according to the *consensus gentium* it is, in fact, morally permissible.

"You call to him to get out of the way, but he is unable to hear you. Is it permissible to light the explosives, blowing through the barrier—and killing your colleague in the process?"

“Certainly it cannot be that once a man has committed any crime, thus becoming ‘guilty’, he can be punished in any way—even if such punishments would save many innocents.”

The Gunman scenario indicates that we must again revise our agent-relative principle; this time, however, it is not so clear what about the action makes it permissible. One interpretation seems to be that somehow the five innocents whom the gunman will kill have a stronger claim on their stringent right to life than does the gunman. How can this be? Again, an extensive discussion of rights theories would only serve to divert attention from the issue. However, it is fair to say that because the gunman is [we have good reason to believe] about to infringe upon the stringent rights of other members of our society, he in doing so weakens his claim on the enforcement of his rights by society and its citizens. This is not to say that he waives his right to life by intending to take the lives of others; rather, he no longer merits the equal protection and enforcement of his right to life.

TO PHRASE THIS IN THOMSON’S WORDS, HIS RIGHT to life loses its status as a stringent right *relative* to the right to life possessed by the innocents he intends to kill. As this is painfully vague, perhaps an analogy will cast some light upon the confusion. Let us consider another variant of the Gunman story: A man walks up to you and tells you that he will kill five people today unless you kill him first. However, rather than recognizing him as a raving maniac, you recognize him as the newly hired state executioner, and you are aware that there are five executions scheduled for the day. Is it still moral to kill this gunman (the executioner)? It is does not appear so; however, the only difference is that the gunman is transgressing societal rules and his victims have not done so, whereas the executioner commits no transgression and his victims have. Thus, the gunman has less of a claim on his rights than do his victims, while the executioner has more of a claim on his rights than do his victims—and we may shoot the gunman, but not the executioner. It is clear that NNT’ in no way addresses the concept of relative innocence and guilt of the agent compared to the victims; this appears to be the shortcoming of NNT’. From this we can arrive at a further refinement of our principle, which now posits:

One may not create a new threat which is intended to inflict harm upon an innocent party as an end in itself or as a means to saving the many; one may not create a new threat which inflicts intentional harm upon a guilty party unless in doing so one prevents intentional harm from being inflicted upon an innocent party; and one may not divert an existing threat from a larger party to a smaller party in such a way as to grossly impinge upon the stringent rights upon which the smaller party has a claim.

Before we proceed, let us make more clear what is meant by certain terms in this re-revision of Thomson’s principle. “The many” is intended to indicate a number of people either greater than or equal to that of the “innocent party”. The definitions of “innocent” and “guilty” are somewhat more complex. Certainly it cannot be that once a man has committed any crime, thus becoming “guilty”, he can be punished in any way—even if such punishments would save many innocents. For example, it is disturbing to imagine taking a thief out of prison, presenting him in the city market, announcing that any murderers will receive his fate, and then shooting him. Thus, it must be that the guilt and the threat created must be intrinsically linked. In other words, the threat must be brought about as a specific response to the action(s) which bring us to label the agent guilty. All of the other terms are assumed to have the same meanings as they have throughout our discussion. This rephrasing of the NNT’, which we will call NNT’’, does seem to account for our sentiments in the Gunman scenario, for we are not creating a new threat which inflicts intentional harm upon an innocent—as we have good reason to believe that the gunman will in

fact kill the five innocents, he no longer posses “innocent” status. This is the purpose of the rephrasing of the first clause in the principle; the new second clause serves as a check upon the abuse of guilty parties. For instance, let us imagine that the gunman in the scenario never approached us, but we happened to recognize him walking down the street. Would it be permissible, even if you were completely sure that it was in fact the gunman, to shoot him? I do not believe that there is anyone who would respond yes, it is morally permissible to do so. However, were the first clause to stand alone, as he is not an innocent party, he would have no claim upon our abstention from harming him. The second clause prohibits actions against guilty parties which are not intended to immediately protect innocent parties. This coheres with our brief discussion of rights, because the guilty party does not completely cede his rights; rather, he simply has less claim on them than do his innocent victims. Thus, NNT'' seems to be the necessary modification of NNT', as it correctly predicts our relative sympathies when confronted with parties of different moral and legal statuses.

Even this second revision, however, is not intended to be the ultimate expression of our moral intuitions. Rather, it is obvious upon careful introspection that there are cases in which the principle as it stands is inadequate. This is not proof that the principle I have offered is invalid. All this indicates is that I have not reached the ultimate revision of Thomson's principle. Were this second revision to be revised, and revised again, *ad infinitum*, the result would undoubtedly be a principle which captures the entirety of our moral intuitions. It is possible that such an ultimate principle would in fact be rather simplistic; it strikes me, however, that such a complete formula would necessarily be as complex as the brain which applies it. Nonetheless, such questions must be laid aside and returned to another time.

IT IS RELATIVELY CLEAR, HOPEFULLY, THAT THE final expression of the agent-relative principle I have arrived at in this work, NNT'', is able to predict the majority of our reactions in hypothetical scenarios. For such a principle to have true practical value, however, it should be able to explain and predict actual case-studies as well. One interesting test comes from a recent news event. An ardent supporter of the Right to Life (the movement which finds abortion immoral because it is murder of the fetus) killed an abortion clinic doctor. Upon being asked for her justification for her actions, she stated (and some agree with her) that her action was moral because it prevented the murder of innocent fetuses. Understanding her opinions and beliefs, her action coheres to NNT'', because while she did create a new threat which intentionally harmed somebody, in her mind her victim was a guilty party—a murderer, one who murdered every day. Because of this, it was morally permissible for her to kill the doctor, because in doing so she prevented intentional harm from being inflicted on innocent parties—unborn fetuses. Thus, understanding her intuitions, it should come as no surprise to us that she found her action morally permissible. Equally predictable is the fact that proponents of the legality of abortion find her action morally impermissible, as the doctor—an innocent party in their eyes—had no lessened claim on his stringent right to life. Because he was not a guilty party, there could be no justification of subordinating his rights to the rights of the fetuses, even if it were granted that such rights exist.

Thus, NNT'' does in fact accurately predict real-world reactions and opinions, even in cases in which individual intuitions differ radically. It is this applicability to everyday events which demonstrates the success that this new agent-relative principle has in explaining difficult situations. ◻

“The guilty party does not completely cede his rights; rather, he simply has less claim on them than do his innocent victims.”
