

compos mentis

Prima Facie Consequentialism: Reconciling Consequentialism and Deontology into a Normative Ethical Theory

Logan B. Cross

Michigan State University

BIOGRAPHY

Logan Bradley Cross studies philosophy and psychology at Michigan State University where he serves as a lab instructor for an undergraduate course in ancient Greek philosophy and a philosophy tutor. His primary areas of studies include ethics, political philosophy, sociology, aesthetics, Eastern philosophy, and critical theory. He was born in Muskegon, Michigan.

ACKNOWLEDGMENTS

Thank you to Professor Matthew Ferkany whose teachings would inspire this work. Furthermore, thank you to the attendants of the 2018 Michigan Undergraduate Philosophy Conference hosted by the University of Michigan Flint and the *Compos Mentis* team for allowing me to share and test my ideas; their wise observations would lead to the third section of this work.

ABSTRACT

Deontological ethics is defined as a normative ethical theory that is based primarily around rights, duties, and/or obligations that *always* take precedence in moral judgements even if following said duties would not lead to the best consequences. For example, many deontological philosophers view abstaining from murder as a deontological duty; to such thinkers, it is *always* wrong to commit murder, no matter the consequences. The rationality of deontological ethics has been hotly debated and has a propensity for dividing opinions. In the first section of this essay, I will argue that deontological ethical theories are irrational because they are products of innate, automatic cognitive heuristics loosely tied to human emotion as opposed to genuine moral reasoning; in the second section, I will try to reconcile deontological judgements into a working normative ethical theory by suggesting that they have value as *prima facie* duties—or duties to be observed under all circumstances except where carrying out said duty would have foreseeable negative consequences for the collective whole, or another *prima facie* duty that is contradictory to another is in that instant more important; in the third section, I will aim to distinguish *prima facie* consequentialism from traditional utilitarianism, using examples to show how the two doctrines differ in addition to how *prima facie* consequentialism is capable of avoiding the criticisms typically associate with traditional utilitarianism.

KEYWORDS

Prima Facie Consequentialism, Ethics, Ethical Theory, Consequentialism, Deontology, Morality, Utilitarianism, Cognitive Heuristics, Moral Cognition, Normative Ethical Theory

PART I: MORAL AUTOMATION—A CRITIQUE OF DEONTOLOGICAL ETHICAL JUDGEMENTS

The theory that deontological theories are grounded in emotion rather than reason was discussed in detail by philosopher Joshua Greene in his essay *The Secret Joke of Kant's Soul*. Greene makes use of two famous ethical dilemmas; namely, the *trolley dilemma* and the *footbridge dilemma*. The trolley dilemma goes as follows; a runaway train is poised to hit and kill five people, unless you throw a switch that will divert the train onto a separate set of tracks where one other person is standing, thus killing the one individual instead five. The footbridge dilemma is of a similar nature— as with the trolley dilemma, a runaway train is poised to hit five people standing down the tracks. However, in the footbridge dilemma, rather than throwing a switch to divert the train, you can push a heavy person down from an overlooking footbridge onto the tracks to stop the train. Surveys and other testing has shown that people tend to be more willing to throw the switch in the trolley dilemma than to push the person onto the tracks in the footbridge dilemma even though both situations have the same cost/reward balance—killing one person to save five others (Greene 2015, 42). Kantians¹ will likely attribute this fact to the *doctrine of double effect*²; however, Greene dismisses this notion by pointing out that the people tend to react similarly to the trolley case in the *looped trolley dilemma*, in which the runaway trolley is poised to hit five people unless you throw the switch towards an alternate *looped* track with one person poised to be hit that would stop the train; however, in this case the train would have looped back around and hit the five had the one person not been hit. Greene points to the fact that in both the footbridge and lopped trolley case a person is directly used as a means, saying “[t]he consensus here [in the looped trolley dilemma] is that it is morally acceptable to turn the trolley in this case, despite the fact that here, as in the *footbridge* case, a person will be used as a means” (Greene 2015, 42). As such, Greene proposes a different theory to account for the difference—human emotion; Greene feels that the only reason people sacrifice the one person in the trolley dilemma and not in the footbridge dilemma is because the act of physically

-
1. Philosopher Immanuel Kant was the most accomplished proponent of deontology as a normative ethical system.
 2. The doctrine of double effect states that is morally permissible to cause something negative to happen as a side effect of causing a good so long as it was unintended. This implies that it is wrong to use a person directly as a means to prevent a negative occurrence.

pushing a person to stop a train is more emotionally traumatizing than merely throwing a switch. As evidence, Greene references previous scientific studies that show that the parts of the brain associated with emotion, rather than cognitive reasoning, are activated when people make deontological judgements.

A deontologist, of course, will be unconvinced by Greene's theory. They will stand by Kant, who proposed deontological judgements as a normative ethical system, saying "[an] action's moral value doesn't depend on whether what is aimed at in it is actually achieved, but solely on the principle of *the will* from which the action is done, irrespective of anything the faculty of *desire* may be aiming at" (Kant [1785] 2008, 10). So, Kant feels as though acts such as throwing the switch are wrong simply because it is always wrong (that is, it is a deontological duty not) to kill somebody.

In his essay *Morally Irrelevant Factors: What's Left of the Dual Process-Model of Moral Cognition?* Philosopher and deontological thinker Hanno Sauer discusses several logical and method flaws in Greene's theory. One flaw concerns the induction power of Greene's form of inference from which he draws his conclusion, which Sauer identifies as *reverse inference* (Sauer 2012, 789). The form of a reverse inference is as follows: (1) Whenever cognitive task T is performed, brain area A is active. (2) Other studies have shown that whenever cognitive task T is performed, brain area A is active. (3) T is a kind of K. (4) Activation of A while performing T demonstrates that T is of kind K. This form of inference differs from the form predominantly used in cognitive neuroscience, the *forward inference*³. According to Sauer, the reverse inference is not immediately clear or convincing like a forward inference is. He notes that the strength of the reverse inference is directly correlated with the selectivity of activities the activated brain region seems to be involved in, saying "[t]he strength of a reverse inference varies as a function of the selectivity with which the region is involved in the process" (Sauer 2012, 789). In other words, if the brain areas Greene cited in his studies were for the most part only observed in dealing with emotional tasks, Greene's theory might be cogent. However, Sauer suggests that the brain areas Greene based his studies on do not appear to have this selective quality, as the brain regions Greene cites seem to play vital roles beyond emotion. Sauer even suggests that the primary functions of these brain areas do not appear to be related to emotion, saying "[t]

3. (1) Whenever cognitive task T is performed, brain area A is active. (2) Therefore, brain area A is involved in the execution of T.

he primary function of the dlPFC [a brain area cited by Greene] seems to lie in working memory (Wager & Smith 2003): an increased activation in that area does not directly bear on whether a mental function is more cognitive, in the sense of rationality, but on whether it manipulates a greater amount of information" (Sauer 2012, 790).

In defense of Greene's theory, Sauer's evidence that the primary functions of the brain areas Greene cited in his work does not prove that said brain areas are not still related in some way to emotions present in deontological judgements or emotions in general—it is not necessarily unlikely, as brain areas have been known to have multiple/complex functions. Furthermore, Sauer's objections to Greene's physical evidence does not show deductively that emotions do not play a role in deontological judgements in general, nor does it give any insight to the question of whether emotions are truly a morally irrelevant factor. It seems safe to infer from the very natures of emotion and cognitive reason that, so long as one has ample time to reason, one should exercise reason in all ethical judgements. That is, emotion is often associated with deluded and rash decisions⁴, whereas reason and thoughtful reflection have always been associated with wisdom. Considering this, it seems cogent to say that cognitive reflection should always take precedence over emotionality in ethical judgements. This, of course, suggests Greene is right to think that emotions are morally irrelevant factors in ethical judgements.

Sauer remains unconvinced. He believes that the link Greene draws between matters of automatic versus controlled mental processes, emotional versus cognitive processes, and justified versus unjustified processes of treating ethical issues suffers from *conflation*⁵ (Sauer 2012, 800-801). To Sauer, emotionality and automatic (as opposed to cognitive/reasoned) brain processes need not be merged together, saying "[a]utomatic processes need not be emotional at all: non-emotional, but nevertheless quick and effortless intuitions in logic, language or physics are apt counterexamples to this assumption" (Sauer 2012, 801). Sauer goes on to suggest that the automatic responses involved in deontological judgements might be sort of moral *heuristics* that can be vital to our functioning.

4. In other words, emotionality seems to be tied to decisions the brain makes *automatically*—more on this later.

5. Conflation means to merge two or more idea/belief sets into one—in this case, Sauer implies that Greene's conflation of emotionality, automatic cognition and deontological judgements is unwarranted.

Heuristics are general rules/procedures followed to solve a specific problem. In the case of cognition and neuroscience, heuristics often refer to automatic cognitive responses to certain stimuli; fears and biases that are registered automatically in people more or less without their conscious control are examples of these. Greene, like many neuroscientists and evolutionary psychologists, even suggests that such heuristics might have an evolutionary origin; more on this to follow (Sauer 2012, 795).

Greene's view of moral heuristics represents a relatively modern evolutionary psychology view of the brain and its functions. Followers of evolutionary psychology believe that the automatic processes of the brain (the so-called cognitive heuristics Sauer discusses) are evolutionary relics of the mind. For example, consider that many people have a strong fear of heights that arises in them automatically—an evolutionary psychologist would account for this by saying that our brains evolved an automatic response towards heights via natural selection because among our distant ancestors those who did not fear heights were more likely to fall and die from heights and thus not live on to reproduce relative to those who feared heights. Greene feels that these automatic deontological responses present in humans may be of a similar evolutionary nature—sort of moral heuristics that have been passed down through evolution because they were beneficial to the survival of our ancestors; it is likely that they evolved so that people would react in time to situations that required *urgent* ethical decision making—for example, one might not have time to think through options if one wishes to save a person from drowning, as the individual might drown before one has time to amply reason and weight the risks, etc.; the situation needs a quick, automatic decision. Here we find the strength of these so-called deontological judgements that appear as evolutionary relics; they create automatic responses to these urgent situations. As such, Sauer can suggest that the automatic deontological judgements that Greene dismisses as emotional and unreasoned may, on the contrary, be of significant, evolutionarily-proved, time-tested value.

Returning to the defense of Greene, I suggest that the jump from a deontological judgement being an evolved automatic cognitive response to deontological judgements truly having moral value is guilty of committing the *naturalistic fallacy*⁶. To illustrate this view, I cite modern evolutionary psychological

6. The naturalistic fallacy is an informal fallacy that is committed by trying to infer that something is good/we ought to do something from what is natural/the way something is.

findings regarding discrimination. Evolutionary psychology has shown empirically that many people have an *implicit bias*⁷ against people of different race/ethnicity than their own, often leading to racist and/or discriminatory behaviors. Evolutionary psychologists believe this automatic response is an evolved automatic response that came about because it was beneficial for the early ancestors to distinguish between coalitional boundaries for personal safety (which often meant identifying cues associated with physical similarities). While these discriminatory behaviors are a natural, automatic response much like Sauer's view of automatic responses associated with deontological judgements, very few people would advocate the notion that racist, xenophobic, or discriminatory behavior is morally justified. In fact, many people explicitly reject such behaviors. From this, the presence of the naturalistic fallacy seems clear—from the same principle that says that you cannot derive the notion that we ought to act with discrimination because we have an automatic and natural tendency to do so, one cannot deductively say that we ought to claim automatic deontological responses are morally valuable simply because we have a natural tendency to make such responses. The Stanford Encyclopedia of Philosophy summarizes the naturalistic fallacy by saying “[t]he intuitive idea [of the naturalistic fallacy] is that evaluative conclusions require at least one evaluative premise—purely factual premises about the naturalistic features of things do not entail or even support evaluative conclusions” (Ridge 2017). Sauer's argument lacks this evaluative premise—while he does give good support to the notion that deontological judgements stem from cognitive heuristics, he fails to provide any evaluative support that such heuristics are morally beneficial. While it is true that Sauer only suggests that deontological emotions only *usually* respond to morally relevant features of situations, Sauer cannot hope to create a normative ethical system that will *always* respond to morally relevant factors based around said cognitive heuristics due to this naturalistic fallacy. At best, deontological judgements represent *prima facie* duties⁸, as described by Ross (Ross 1930, 21).

In conclusion, I attest that while Sauer's views have a certain degree in power in discounting the physical evidence that Greene cites, it does not show deductively that deontological judgements are emotion-free altogether, nor does it provide

7. Implicit biases are subdued judgements or behaviors that result from sub-conscious cognitive processes.

8. That is, conditional duties that are to be followed in most circumstances, but that can be overridden if the situation requires it.

insight into the matter of whether emotionality and the automatic cognitive heuristics truly are morally irrelevant factors. As such, I conclude that deontological judgements are at best evolutionary cognitive heuristics that evolved so humans could quickly make moral judgements in situations where time pressures do not allow for further reasoned ethical consideration. While such deontological judgements may *usually* be reliable, it seems apparent from both a normative and empirical standpoint that they are not *always* reliable. Considering this, it seems cogent to say that Greene's theory of the moral irrelevance of deontological judgements is correct, and that we should always, so long as we have ample time to do so, attempt to reason through moral judgements in the cognitive, non-emotional way that Greene finds more so characteristic of consequentialist as opposed to deontological judgements.

PART II: THE CONSEQUENTIALIST SOLUTION

We have seen in part one that Consequentialism can be defined as a normative ethical theory that states that an act is morally right if it produces the best overall consequences, regardless of the actions taken to achieve the best consequences. By definition, it follows that consequentialism allows one to occasionally do negative acts as an unintended side effect of creating the best outcome. This ethical theory stands in opposition to deontological ethics, which states that there are some moral duties to be held regardless of intentions or consequences. In this essay, I will argue that consequentialism is the most rational foundation for a normative ethical system and that, furthermore, deontological judgements only have value as moral rules-of-thumb called *prima facie duties*; I will then use this notion of *prima facie* duties to reconcile consequentialist and deontological moral philosophies into a normative ethical system that is both convenient and reliable, which I term *prima facie consequentialism*. Finally, I will conclude this essay by defending *prima facie* consequentialism against some anticipated objections that one may raise.

In effort to show that consequentialism is the most rational model for a normative ethical system, I would first like to utilize an argument from Aristotle's *Nicomachean Ethics*. Aristotle claims that in order to understand the human good, we must understand the purpose of the good. He writes "[s]urely it is that for the sake of which the other things are done; and in medicine this is health, in generalship victory, in housebuilding a house, in another case something else,

but in every action and decision it is the end, since it is for the sake of the end that everyone does the other things" (Aristotle [1995], *Nicomachean Ethics* I.7 1097a19-23). By this, Aristotle means to say that to know if something is good/effective, we must know its function; its end in its teleology, in other words. The rationale behind this is that we can then know if something is good by measuring how well it achieves that end; if it successfully achieves its desired function it is good, whereas it is not good if it fails to achieve its desired function. For example, one would not be able to make a statement on the goodness of a clock unless one understands its teleological aim—namely, to accurately tell time. But if one knows this goal, one becomes entitled to rationally say if that particular clock is good, based on if it performed its function well by consistently keeping accurate time. If one accepts this notion, one would also likely accept the closely related statement: "*to understand the moral good, one must know the function of morality.*" If this statement is true, then it seems to have implications for the rationality of a normative ethical system. Namely, if a normative ethical system is rational, it must respond to the function/reason of morality—the 'why' behind moral actions.

The question of why we ought to act morally is debatable, though one can still speculate and expect relative accuracy, as it seems clear that the purpose of morality seems to be tied more or less to *goodness*⁹. One answer that one might first arrive at is that we ought to be moral because it is good for the people as a collective whole. If this is true, then what is most moral is none other than that which is most beneficial to the collective whole; that which produces the best *consequences* for the collective whole. This is completely consistent with consequentialism. Deontologists, however, seem to require alternate reason(s) for the purpose of morality because deontologists suggest that there are moral duties that are to be observed *even if* breaking said duties were to produce the best consequences. For example, if one is a deontologist who believes that abstaining from murder is a moral duty, then one could not possibly justify murder, even if the situation were such that it would produce the best consequences (such as murdering a dictator before he causes the death of countless others). As such the deontologist cannot say that the purpose of morality is that we ought to be

9. By *goodness* I mean the flourishing of the collective whole. It should be noted that this collective whole implies *everyone*, not merely the *majority* of people—this disclaimer, I hope, protects this theory from certain criticisms.

moral because it benefits the collective whole—for the deontologist, there must be something more to morality's purpose. What this addition could be is not immediately clear, and until such an addition can be identified, this seems to suggest that consequentialism is the more rational theory to be base a normative ethical system upon.

In light of this, any deontological ethical system that does not simply state that one should always do what produces the best consequences for the collective whole is likely responding to some morally irrelevant factor(s). If this is true, then the irrelevant factor(s) should appear in situations where deontologists opt to take the action contrary to what produces the best overall consequences for the collective whole for some reason. This can be further examined via the trolley dilemma mentioned earlier. The consequentialist answer to this dilemma is to throw the switch because the result of one person dying is better than the five dying. The majority of people, including many deontologists, agree with throwing the switch in this situation. However, the trolley dilemma has numerous related dilemmas where interesting differences arise between the consequentialist and deontologist beliefs.

One such related situation is the footbridge dilemma, mentioned in part I; recall that this situation is usually met with more controversy. Many deontologists say that it is wrong to push the heavy person down from the tracks even though it would again produce the best consequences in that only one person dies as opposed to five because of certain moral duties that transcend any intentions or consequences. Immanuel Kant is perhaps the most accomplished and noted philosopher to take this view; he writes “[s]o an action’s moral value doesn’t lie in the effect that is expected from it, or in any principle of action that motivates it because of this expected effect. All the expected effects—something agreeable for me, or even happiness for others, could be brought about through other causes and don’t need the will of a rational being, whereas the highest good—what is unconditionally good—can only be found in such a will. So this wonderful good, which we call moral goodness, can’t consist of anything but *the thought of law in itself that only a rational being can have*—with the will being moved to act by this thought and not by the hoped-for effect of the action” (Kant 2008, 10). Kant clearly believes that moral laws/duties can be found through reasoning, and these duties ought to be followed under *any* circumstances.

Kantians can use such duties to help to explain the differences exhibited in the trolley dilemma as opposed to the footbridge dilemma. One duty commonly cited by Kantians to help explain that very distinction is the doctrine of double effect. Recall that the doctrine of double effect states that it is morally permissible to cause a negative effect as a foreseeable but unintended side effect of trying to cause some greater good. Applied to the dilemmas, this means that it is permissible to throw the switch in the original trolley dilemma because the death of the one bystander, though foreseeable, was unintended. The death of the heavy man in the footbridge dilemma is intended, and therefore immoral. By this doctrine, inflicting harm is better than simply allowing it; or, in technical terms, it is wrong to use somebody as a mere means, as the heavy man was used in the footbridge dilemma.

Although the doctrine of double effect may seem to be a reasonable explanation for the differences exhibited in the trolley dilemma as compared to the footbridge dilemma, if the looped trolley dilemma would be examined (recall that the looped trolley dilemma states that a runaway trolley is poised to hit five people, and, again, you can throw a switch which will send the trolley to an alternate set of tracks where one bystander is standing who will be killed as a result. However, in this case, the alternate set of tracks is looped back to the main set of tracks, so that if the trolley did not hit the one bystander and thus slow down, the trolley would return to the main set of tracks and again hit the five people). Because the death of the bystander in this case is necessary to save the five, as with the footbridge dilemma, it follows that the bystander is again used as a mere means. And yet, many people, deontologists included, believe that it is morally acceptable to pull the switch to save the five in the looped trolley case. As such, the idea that the doctrine of double effect can explain differences in general reactions to the trolley dilemma as opposed to the footbridge dilemma becomes doubtful.

This notion allows two questions to remain; what, then, explains the differences in reactions to the trolley versus the footbridge dilemma, and does this reason respond to morally relevant factors? Recall that these questions were discussed by Greene in *The Secret Joke of Kant's Soul*; he writes "[t]he consensus here [in the looped trolley dilemma] is that it is morally acceptable to turn the trolley in this case, despite the fact that here, as in the footbridge case, a person will be used as a means" (Greene 2015, 42). With the doctrine of double effect thus rejected,

Greene then offers an alternative explanation to the differences observed in reactions between the trolley and footbridge dilemmas; emotion. He claims that “deontological judgements tend to be driven by emotional responses, and that deontological philosophy, rather than being grounded in moral *reasoning*, is to a large extent an exercise in moral *rationalization*” (Greene 2015, 36). To apply this to the family of dilemmas mentioned earlier, Greene would say that the reason so many are willing to throw the switch in the trolley dilemma but not to push the heavy man down from the bridge in the footbridge dilemma is simply the result of unconscious emotions; physically pushing a person down from the bridge is much more emotionally traumatic than simply throwing a switch, and it is this fact that accounts for the differences in reactions between the trolley and footbridge dilemmas.

If this is so, it seems that deontological judgements are responding to factors that are morally irrelevant, or—if indeed they are at times reliable—they are prone to misfire, though this may not necessarily be the case. That is, it seems that one could reasonably argue that our emotions often seem to pique naturally in response to things we genuinely feel are wrong and would defend as such on a rational level, and thus emotions may actually respond to morally relevant factors. For example, one might feel sad or angry if they witness a person being robbed; and understandably so, as few would argue that stealing is generally immoral, aside from in some extreme cases. From this fact, one can conclude that emotions might have a morally relevant role in one’s life as a sort of guide that hints at something’s goodness or badness. A normative ethical system based on deontological obligations would therefore be reasonable, as even if deontological intuitions of obligation are largely responding to emotion, they would still be responding to something seemingly morally relevant.

From this it seems that deontology is still a reasonable approach to creating a normative ethical system; however, because of the volatile nature of emotions, I argue that this is not so. Nobody will argue with the notion that our emotions often lead us to poor decisions; an angry person may, for example, punch something in anger, seldom leading to anything more than an injured hand. This example shows that emotions are at very least unreliable guides for making judgements. Furthermore, if emotions can lead us astray in our everyday decision making, as seems to be the case, it is cogent to think that they can lead us astray morally, as well. Indeed, this seems to be the case, as illustrated previously by the distinction

between the footbridge and looped trolley dilemmas. Deontology, then, is certainly not an infallible approach to morality and if one wishes to create an infallible normative ethical system—the sort of universal template to determine if something is morally right that Plato’s Socrates¹⁰ and other philosophers have so vigorously searched for—then deontology will not do.

Consequentialism, however, does not suffer from this shortcoming. If it is true that the purpose of morality is simply to produce the best consequences for the collective whole as I have suggested earlier in this essay, then it seems a normative ethical system that is solely based on doing whatever produces the best consequences for the collective whole could never fail or misfire such as a system based on emotion. It is ultimately for this reason that I assert that consequentialism is the most rational basis for the creation of a normative ethical system.

But there is still something left unsaid for the value of deontological judgements; some deontological judgements might lead to correct moral decision in almost every case—murder, for example, can be thought to be immoral in *almost* every conceivable situation. Surely, then, one ought not to assume that deontological judgements are completely devoid of value; they may even hold a considerable amount of value, despite their apparent imperfection. This is clear even when looked under the lens of Aristotle’s function argument discussed earlier in this essay—to again take the clock example used previously, we would not say that a clock is not a good one for losing a second or two of time each year. As such, we should not try to claim that deontological judgements are not good merely because they have the minor imperfection of possibly leading us astray morally in certain extreme, isolated scenarios. Simply put, actions performed based off deontological judgements will *almost* always be morally correct, but there are extreme situations that render deontology ineffective. They therefore have value not as the absolute duties that Kant defended, but rather as what philosopher W.D. Ross called *prima facie* duties. I believe that philosopher Russ Shafer-Landau gave the best definition of *prima facie* duties in a commentary on an excerpt from

10. It is apparent that Plato—and perhaps Socrates—were interested in a universal template for determining the goodness of moral judgements, as in Plato’s *Euthyphro* dialogue Socrates inquires of Euthyphro to give him a model of which to judge the morality of actions, saying “[t]ell me then, what this form [of moral goodness] is so that I may look upon it, and using it is a model, say that any actions of yours or another’s that is of that kind is pious, and if it is not that it is not” (Plato [2009], 6).

Ross in his *Ethical Life*, defining a *prima facie* duty as “an always-important reason that generates an ‘all-things-considered’ duty, provided that no other reason or set of reasons is weightier in the situation” (Shafer-Landau 2015, 126). Under this definition, deontological judgements can have value, though they have what can be seen as a failing in that they are not universal—situations can arise in which the deontological judgements no longer hold normative value.

It is precisely this failing that consequentialist thinking patches, by acting as a filter for the flaws in moral reasoning that our emotions sometimes (but not always or perhaps even often) create. For example, a deontologist may claim that not killing is a moral obligation. This seems cogent, as most would agree that killing is *almost* always wrong. However, this same majority will often concede certain situations where extraneous factors seem to allow for killing, such as killing a dictator before said dictator causes death and suffering to those under his rule. This is nothing more than consequentialist principles testing deontological judgements and correcting for the extremity of the situation; a correction which seems to be based on nothing more than trying to create the best consequences for the collective whole. I term this normative ethical system of consequentialist-filtered deontological judgements *prima facie consequentialism* and will refer to it as such for the remainder of this essay.

As with any approach to ethics, this thesis is subject to criticism. One criticism that can be anticipated for *prima facie* consequentialism is that it may simply be a form of deontology in disguise. It may be the case that deontology is still applicable to the case of choosing to kill the dictator despite there being an obligation to avoid killing others if some other deontological duty so-to-speak superseded the other duty. To illustrate this in the context of the dictator example used previously, it may be the case that, although there is a deontological obligation to avoid killing another person, there may be another deontological duty to protect the well-being of other people that supersedes and nullifies the obligation to avoid killing another should these two deontological duties ever come into conflict. One can take this as suggesting that deontology cannot be so quickly dismissed.

While this worry certainly has merit, I believe that *prima facie* consequentialism can be derived and perhaps even strengthened by this concern. This is because if it is so that two deontological obligations can conflict with each other and one is *determined* to be of greater importance than the other and is thus carried out rather than the other, then unless the reasoning behind why one obligation should

be favored over the other is completely arbitrary¹¹, it follows that there must be a *determining factor* behind why that obligation was favored over the other obligation in the conflict¹². The most rational notion reason that could be given for such a favoring seems to be that in that situation, favoring that obligation as opposed to the other creates the best consequences for the collective whole; the obligation to protect the lives of the people outweighs the obligation to avoid murder because doing would be best for the people, who are the majority. If this is so, it seems that consequentialism underlies our decision making for what obligations one ought to follow should one find themselves in an extreme situation where the *prima facie* duty of deontology come into conflict with one another. In other words, consequentialism provides us with the criteria to rank the importance of deontological duties if they happen to conflict. This fact, coupled with the fact that one could doubt that the so-called obligations of deontology could even be able to be considered proper duty since situations can arise where they are nullified for the greater good, seem to credit *prima facie* consequentialism as true. A proper duty, that is, would seem to demand a sort of absoluteness that is missing from deontology. If the deontologist wishes to defend true deontology, they must find a duty that applies in every situation and circumstance without fail, one that is never irrelevant. But such a duty, it seems, cannot be found, for no matter how important of a duty one believes a duty to be, one can seemingly always think of an example of a situation where it might be best to forsake that duty in the name of producing the best consequences for the collective whole. As such, *prima facie consequentialism* remains plausible.

Another likely objection to *prima facie* consequentialism can be derived from another common deontologist objection to consequentialism; that is, that consequentialism is devoid of human emotion, which deontologists feel is an important foundation for morality. That is, they feel that consequentialism treats the process of making a moral judgement as nothing more than a sort of moral equation to be solved, which must abandon emotionally relevant aspects of life such as familial ties, friendship, and other aspects of life where one might have a duty

11. A premise that I feel most will surely wish to avoid.

12. This line of reasoning is again reminiscent of Aristotle's function argument of human goodness discussed earlier in this essay; that is, if one *prima facie* duty is determined to be more important than another in a certain situation, then it follows that there must be a purpose of why that duty was chosen relating to some teleological purpose—I argue that this purpose is to create the best consequences for the collective whole, the central premise of consequentialist thinking.

to be partial for emotional and trust-related reasons. The deontologist may claim, for example, that throwing the switch in the trolley dilemma abandons a certain aspect of humanity because it is done merely for the sake of saving more people; humanity, some deontologists go on to claim, must be much more complicated and emotionally deep than the mere numbers game the consequentialist appears to support. I, however, argue that this objection is founded on an inaccurate portrayal of the consequentialist in action. Those that object to consequentialism along these lines seem to see consequentialists as cold, unfeeling calculators who would willingly kill off forty-nine percent of the population if it would save the other fifty-one percent and would do so without reluctance or regret.

In reality, this is simply not the case; there is nothing that says that a consequentialist must be unfeeling about a *prima facie* duty broken in lieu of protecting a more consequentially relevant *prima facie* duty. Ross himself touches on this subject using the example of somebody breaking a promise for the greater good in his *What Makes Right Acts Right?* He writes “[w]hen we think ourselves justified in breaking, and indeed morally obliged to break, a promise in order to relieve some one’s distress, we do not for a moment cease to recognize a *prima facie* duty to keep our promise, and this leads us to feel, not indeed shame or repentance, but certainly compunction, for behaving as we do” (Ross 1930, 21). From this, it is plain to see that the consequentialist has no obligation to be emotionally devoid in making complicated moral judgements. In fact, the consequentialist is bound to feel a certain sense of loss if they were required to violate one *prima facie* duty for the sake of another—it is merely that the consequentialist knows that one must not allow said emotions to control their moral judgements, for they can mislead one if not filtered by reason.

Perhaps a more serious allegation against *prima facie* consequentialism that could be anticipated is a situation in which doing what seems to produce the best consequence for the collective whole requires doing an act so sickening that it seemingly could not be justified even if it did produce the best overall consequences. A commonly cited example that this can be seen in involves the *backpacker dilemma*, which is yet another sister dilemma to the trolley dilemma discussed earlier in this essay. In this dilemma, five people are in the hospital in dire need of an organ transplant to stay alive. As it happens, there is a backpacker with minor injuries in the hospital whose organs, if harvested, could save those five at the expense of his single life; however, the backpacker has no intention

of willingly relinquishing his organs for this cause, and they would have to be harvested secretly against his will as such. Note that this situation presents the same life/death ratio presented in each of the trolley dilemmas—namely, five people will survive at the expense of one death. As such, many people will claim that the consequentialist is forced to believe that the morally correct action is to harvest the backpacker's organs against their will. Many who accept this notion use this as a *reductio ad absurdum* argument against consequentialism, as it certainly seems as though there is something *clearly* immoral about harvesting the organs of the backpacker.

Even I as a consequentialist will admit that harvesting the organs of the backpacker seems very wrong indeed. However, I deny the premise that the consequentialist is forced to believe that harvesting the organs of the backpacker is the morally correct choice¹³. I deny this claim on the grounds that it may be a misinterpretation of what choice would actually lead to the greatest overall consequences for the collective whole. It is true that five lives may typically be more valuable than one, but the importance of certain principles may transcend the importance of even insuring that the greatest number lives; furthermore, what *would actually* lead to the greatest overall consequences is not always clear. To relate this to the backpacker dilemma, it may be that if the doctor were to harvest the organs of the backpacker, the trust patients in hospitals have that says that the hospital will do everything in its power to insure the lives of oneself might be negated, which in turn might lead to the consequence of people avoiding hospitals. This mass avoidance of hospitals from the grounds of lack of trust might actually cause more deaths than merely five, as those people who no longer trust the hospitals are prone to die from lack of being able to get necessary medical attention. As such, it can be said that *not* harvesting the organs of the backpacker might actually lead to the greatest overall consequences for the collective whole, even though on the surface of the problem five people die as opposed to only one. Therefore, the backpacker dilemma poses no real threat to *prima facie* consequentialism.

13. Although there is no *obvious* reason to necessarily claim that they shouldn't!

PART III: DISTINGUISHING FACTORS FROM TRADITIONAL UTILITARIAN ETHICS

For *prima facie* consequentialism, one important question still needs to be discussed; namely, is the theory merely utilitarianism in disguise? It could, after all, be said that the principle of always trying to create the best consequences, regardless of the is simply a reformulation of the traditional utilitarian doctrine of acting to create the greatest amount of happiness for the majority. This is an important question, as if *prima facie* consequentialism indeed cannot be distinguished from the traditional utilitarian framework, and as such synonymous with utilitarianism, then it will be subject to the same criticisms of that doctrine—criticisms which, as I believe I will show in the arguments to follow, I feel doom utilitarianism. To be sure, utilitarianism share many similarities and relations, and so this critique must be taken seriously. I believe, however, that *prima facie* consequentialism need not worry. In the following I will argue that *prima facie* consequentialism is distinct from traditional utilitarianism in several important respects which allow the theory to circumvent the problems traditionally associated with the utilitarian doctrine and will highlight these differences through example.

To begin our attempt at distinguishing *prima facie* consequentialism from utilitarianism, we must first come to a definition of utilitarianism. One standardly given definition of utilitarianism is a doctrine in which one ought to do what creates the most pleasure for the greatest number of people. John Stuart Mill, a proponent of utilitarianism, discussed this topic in his essay *Utilitarianism*; he says “[t]he doctrine that is the basis of morals is *utility*, or the greatest happiness principle, which holds that actions are right in that they tend to promote happiness, wrong in proportion as they tend to produce the reverse of happiness” (Mill [1863] 2005, 5). Is this view synonymous with *prima facie* consequentialism? Let us again consider the essence of *prima facie* consequentialism—it is an ethical theory in which our strongly felt moral intuitions serve as general *prima facie*—that is, rule-of-thumb—moral principles that, though reliable in most instances, must be broken in instances where certain two or more such rules cannot be simultaneously upheld because they are contradictory; and, in such cases, we must decide on which of the contradicting *prima facie* rules to uphold by using our best reasoning to try to determine which rule the upholding of which would lead to the best overall consequences for the collective whole. Again, we must ask—are these two ethical systems secretly the same? Upon the first analysis,

the two definitions seem to reveal two immediate distinctions; firstly, *prima facie* consequentialism makes mention of the existence of “rules-of-thumb” whereas utilitarianism does not; secondly, consequentialism uses “causing the best overall consequences for the collective whole” for its guide for deciding actions whereas traditional utilitarianism suggests “causing the greatest amount of happiness for the greatest number of individuals.” Are these differences relevant enough to distinguish the two doctrines? Each must be analyzed in turn.

First, let us consider the difference regarding the addition of “rules-of-thumb” in *prima facie* consequentialism, whereas traditional utilitarianism makes no such references to any such rules. Is this enough to distinguish *prima facie* consequentialism from utilitarianism? Even I, who insists that the two doctrines not be confused for each other, must admit that this does not seem to be a relevant distinguishing factor; after all, a utilitarian thinker could easily incorporate such rules into their own framework, if they so desired. For example, a utilitarian might say that it is a *prima facie* rule to avoid killing another, because such an action is anti-conducive of promoting happiness for the most. However, if somehow there was a situation in which killing a person was necessary to save five others, as in the *trolley dilemma* discussed earlier in this essay, they might say that that action were necessary, as the loss of one individual is less of a blow to the state of happiness in a population than the loss of five would be; if such rules conflicted, they would likely say to choose the one which most maximized the happiness for the greatest number. Asserting the existence of such rules does not in any way damage the utilitarian framework; the traditional utilitarian would likely omit discussions of such rules for the sake of theoretical simplicity, but as shown this need not be the case. Likewise, it seems that I could remove reference to *prima facie* rules and fall back merely on the consequentialist framework underlying the theory.

Why, then, do I feel that the so-called *prima facie* rules ought not be omitted? The answer is that that I feel keeping references to said rules helps to emphasize the point that my theory need not be viewed as an all-out attack against deontological ethical systems as consequentialist and utilitarian systems have often been made out to be. This is because that, while I do claim that such deontological judgements are actually cognitive heuristics that evolved in humans over generations of group-oriented living, and so would be tantamount to a naturalistic fallacy to base an ethical theory around *completely*, I do confess that many commonly proposed deontological duties, regardless of their status as cognitive heuristics, are often

excellent moral guides that save us much time in our moral reasoning, and can even save us from the failure of being “too late” in situations where action is urgently needed, such as swimming off to save a drowning individual. These rules save us valuable time that would have otherwise been spent applying the cognitive process—this seems to be true even on the a subconscious level; we, as individual humans, for example, when we cross paths with another individual in the street or some other public place, do not need to go through a process of reasoning to determine that it would lead to the best consequences for everybody involved to not murder that person where they stand and therefore that we ought not do it. It was obvious from the beginning that consequences of the murder would be worse for everybody than the consequences caused by the refraining from doing so; this knowledge seems obvious to us—practically innate even. In this way, *prima facie* rules are of great value; but when the situations become extreme and the upholding of two or more such rules would be contradictory, some of the rules must be situationally relinquished, and this relinquishing, I argue, ought to be decided by a ranking of *prima facie* rules based on what rules would produce the greatest consequences for the collective whole if upheld, a process that must be grounded in deliberate reasoning. Regardless, the underlying and important point here is that the neither the omission or the inclusion of *prima facie* rules is enough to truly distinguish *prima facie* consequentialism from traditional utilitarianism; we must look elsewhere.

We have seen that the first apparent difference between *prima facie* consequentialism and utilitarianism—namely, reference to *prima facie* moral rules-of-thumb—is not enough to save *prima facie* consequentialism from being synonymous with utilitarianism; this point is more of a matter of having a condensed versus an expounded theory. What then of the second apparent difference we identified between the two doctrines, namely, the fact that *prima facie* consequentialism suggests that we should decide our actions based on what “provides the best overall consequences for the collective whole” whereas traditional utilitarianism suggests that we should do “what provides the maximum amount of happiness for the greatest number of people”—is this difference enough to distinguish *prima facie* consequentialism from utilitarianism? The answer to this question will, of course, depend on if, when we say “creating the best consequences for the collective whole” we really mean the same thing as

when we say “creating the maximum amount of happiness amongst the greatest number of people.”

Well, then, are these statements synonymous? The idea does not seem at all unreasonable to suggest; after all, would not the greatest possible world—the greatest consequence achievable—be the world in which the greatest amount of happiness is garnered for the majority? Furthermore, does not my own ethical inquiry, in suggesting that we ought to search for the purpose of morality to determine what is the moral good—an argument analogous to Aristotle’s argument that we must know the function of the human in order to know what the human good is—sort of suggest that happiness would be the greatest consequence in that Aristotle finds happiness to be the end of reason, the human function, that which is desired for its own sake? All this seems true enough; in many cases, I would say that it is correct that the best overall consequence is the one that provides the greatest amount of happiness to the most people, and it is certainly true that I believe Aristotle’s function argument can help us discover the “why” behind morality. Despite this, I believe that in fact the statements “one ought to do what creates the best overall consequences for the collective whole” and “one ought to do what creates the most happiness for the most amount of people” are in fact *not* synonymous; to see this, each of these questions must be analyzed individually.

In terms about the point about Aristotle, it is true that I utilized his function argument, which declares happiness as the *telos* of the human function; however, in doing so I only intended to utilize the *form* of the argument, stripped of its constituents—just as we can discover if a clock is a good clock by knowing if it does its function—timekeeping—well, and how we can know if a human is a good human by knowing if it performs the human function well, I believe that we can know what the moral good is by analyzing the function of morality, which I believe is creating the best consequences for the collective whole. I did not use this argument to attempt to claim that the function of morality is the same as the function of a human—and in fact I believe there are some important differences between the function of the human and the function of morality, though they may in some ways be related; I believe this will become more apparent in the analysis to follow.

What then of the other claim, which stated that causing the greatest amount of happiness for the most amount of people would be the greatest consequence

an action could possibly attain, therefore revealing consequentialism as utilitarianism, only rephrased? As mentioned earlier, I believe not. To be sure, the attainment of happiness for the majority is likely often a major, and perhaps even often the single most important, consideration one ought to make when trying to act so as to create the best consequences for the collective whole, as *prima facie* consequentialism suggests. It is not, however, the *only* consideration one ought to make when trying to create the best overall consequences for the collective whole. It is here that I believe that the distinguishing factor for *prima facie* consequentialism and traditionally understood utilitarianism lies—in traditional utilitarianism, the *telos* used to determine what the best consequences is something fixed—usually the greatest amount of happiness. This is to say that the *telos* in utilitarianism never changes; this is made explicit by Mill, who said in his *Utilitarianism* “[t]he utilitarian doctrine is that happiness is desirable as an end, and is the only thing that is so; anything else that is desirable is only desirable as a means to that end” (Ross 1930, 24). In *prima facie* consequentialism, on the other hand, we need not be confined to only a single *telos*. *Telos*, in the case of *prima facie* consequentialism, should be looked at as the aim one feels needs to be sought to create the best consequences; it is that which fills in the blank of “one ought to do this or that particular action over another because it creates the best consequences in that it ____.” This is why I say that maximizing happiness for the majority is the single *telos* for utilitarianism—it fills in the blank mentioned above by deeming that “one ought to do this or that particular action over another because it creates the best consequences in that it creates the greatest amount of happiness for the greatest number of people.” Additionally, it seems we ought not confine ourselves to a single *telos*; having a single, fixed *telos* seems to me to be analogous to confining oneself to a single tool for any job despite a full range of tools available in one’s toolbox—and surely this is a problem, as some tools are obviously better suited for some types of jobs than others.

Some examples can help to clarify these points. Imagine, for example, a situation in which a person has five pieces of candy to be shared in any ratio amongst the five people, and each person will be happy so long as they get at least one piece. How should the person divide his candy? In this situation, maximization of happiness amongst the majority seems to be a fine *telos* to determine their actions—in this case, it seems that the person should give each person one piece of candy, as this would lead to the best consequences in that

the most people would end up happy through this action. This *telos*, however, will not hold throughout all situations—again, an example can show this. Imagine a society which is extremely prejudiced against some minority; the Nazi regime’s treatment of the Jewish people might be a true historical example of such a society. Consider, now, the idea that the majority in such a society may derive some sadistic pleasure from the abuse and killing of the persecuted minority; as the majority is the majority, this persecution would seem to maximize happiness in this situation—and yet, something seems clearly wrong here; if one wishes to agree with the statement that persecuting this minority is wrong, then they must concede that, in this instance at least, maximizing happiness is not a viable *telos* to follow in trying to create the best overall consequences for the collective whole. As such, some other *telos* must be found to explain this situation; in the example of the persecuting majority described above, this *telos* might be something like equality—that is to say that, in this situation, the majority ought not abuse the minority even if it would make them, the majority, happy, in that refraining from doing so leads to the best consequences by establishing equality amongst individuals—and, in this case, establishing equality is a more important *telos* than happiness for the majority.

It is clear now that *prima facie* consequentialism is distinguished from traditional utilitarianism in that while utilitarianism typically justifies its actions through only a single *telos*, usually maximization of happiness, whereas *prima facie* offers multiple relevant *teloses*, which can take precedent over one another depending on what the situation calls for. One question, however, still remains—how are we to decide which *telos* we ought to follow in a situation? I believe that the answer to this question must return us to the underlying framework of *prima facie* consequentialism, which suggests that one ought to do whatever leads to the best overall consequences for the collective whole; that is, one ought to attempt to determine all the possible *teloses* that could be relevant in that situation and then rank them according to which *telos*, if followed, would lead to the best overall consequences for the collective whole, noting again that the collective whole is *not* merely the majority of individuals in the world, but rather the summation of every single moral agent in our world. And so, from this it seems the ethical decision-making procedure for *prima facie* consequentialism is a two-part process; first, we must know what end, or *telos*, is the relevant end to follow for creating the best overall consequences for the collective whole. Secondly, one

must determine which *prima facie* rule, if any, is the most effective at pursuing said *telos*. This two-part process is what distinguishes *prima facie* consequentialism from traditional utilitarianism; because traditional utilitarianism only offers us one *telos*, usually maximizing happiness, one need not reason to find the which *telos* is best in that specific situation; the *telos*, maximal happiness, is already decided. As such, utilitarianism is different from *prima facie* consequentialism in that ethical decision-making process because only includes one part—namely, reasoning to determine what action will best uphold the *telos* of maximizing happiness. Because of this two-part as opposed to single-part ethical decision-making process *prima facie* consequentialism is distinguished from traditional utilitarianism, and helps the theory circumvent many of the problems typically associated with traditional utilitarianism, such as the one associated with the example of the majority abusing minorities to obtain happiness.

CONCLUSION

In conclusion, consequentialism is a more reliable basis for a normative ethical system than deontology because it seems to respond to the function of morality—namely, to create the greatest overall consequences for the collective whole—whereas deontology seems to at respond to factors that are at least partially morally irrelevant/fallible, namely emotion. Regardless of this fact, deontological judgements have value in that they seem to *usually* lead one down the correct moral path—it is only in cases where two or more deontological duties conflict that this can fail to be the case. These duties, which are morally relevant until extraneous details of a situation nullify their importance, are known as *prima facie* duties. It is consequentialism that allows us to choose between *prima facie* duties in situations in which they conflict, as we are obligated to choose to fulfill the *prima facie* duty whose fulfillment will create the greatest overall consequences for the collective whole. This principle may sound like it is merely a rephrasing of traditional utilitarian ethics, and therefore subjectable to the heavy criticisms associated with traditional utilitarianism; however, *prima facie* consequentialism should not be confused with utilitarianism in its traditional sense because utilitarianism typically only offers one *telos* to direct us to the greatest overall consequences (usually maximizing happiness), whereas *prima facie* consequentialism allows for numerous possible *teloses* as well as a system to rank them; such an addition allows *prima facie* consequentialism to avoid running into

compos mentis

the criticisms associated with traditional utilitarianism. It is therefore the case that if one wishes to form a normative ethical system that can be universally applied to every moral judgement, one could not do better than the doctrine of *prima facie* consequentialism.

REFERENCES

- Aristotle 1995. *Aristotle: Selections: Nicomachean Ethics*, sec.VI. Translated by Irwin, T., and Fine, G. Indianapolis: Hackett.
- Greene, J. D. 2008. *The Secret Joke of Kant's Soul*. Retrieved February 5, 2017.
- Kant, Immanuel. (1785) 2002. *Groundwork for the Metaphysics of Morals*. Edited and translated by Wood, Allen W. Yale University Press.
- Mill, John S. (1863) 2005. *Utilitarianism*. Edited by Jonathan Bennett.
- Plato, (2009). *Plato: Complete works. Euthyphro*, edited by Cooper, John M. Indianapolis: Hackett.
- Ridge, Michael. 2008. "Moral Non-Naturalism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.
- Ross, W.D. (1930) 2002. *The Right and the Good*. Edited by Stratton-Lake, Philip. New York: Oxford University Press.
- Sauer, H. 2012. "Morally irrelevant factors: What's left of the dual process-model of moral cognition?" *Philosophical Psychology* 25(6): 783-811.
- Safer-Landau, R. 2015. *The Ethical Life: Fundamental Readings in Ethics and Moral Problems*. New York: Oxford University Press.