Spring 2014

The Method of Reflective Equilibrium and the Deontology/Consequentialism Debate

Alexander Schultz
University of Colorado Boulder

Follow this and additional works at: https://scholar.colorado.edu/honr_theses

Recommended Citation
https://scholar.colorado.edu/honr_theses/193

This Thesis is brought to you for free and open access by Honors Program at CU Scholar. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
The Method of Reflective Equilibrium and the Deontology/Consequentialism Debate

Alexander Schultz

University of Colorado
Department of Philosophy
April 8th, 2014

Advisor: Dr. Alastair Norcross, Philosophy

Committee Members: Dr. Alastair Norcross, Philosophy
Dr. Wes Morriston, Philosophy
Dr. David Mapel, Political Science
Definitions of Some Terms:
*Note: these will also be defined as they appear in the project. This guide is intended as a reference.

**Should-E**: ‘Should’ in the epistemic sense. The epistemic norm tells us to seek true beliefs and avoid false ones. Thus, if one ‘should’ do X in the epistemic sense, it means that doing X would help one acquire true beliefs and avoid false ones.

**Should-M**: ‘Should’ in the moral sense. The moral norm tells us to do things that are morally right and to not do things that are morally wrong. Thus, if one ‘should’ do X in the moral sense, it means that it would be morally right for them to do X; this is the same as saying “one ought to do X.”

**iPeer**: Intuiting-peers. Two people are iPeers iff they are both capable of having moral intuitions.

**iPeer Disagreement**: The problem of intuiting-peer disagreement. I argue that we cannot trust a particular moral intuition if it is not the case that a large percentage of people share that intuition, because given the contrary (intuitive divergence) we will be unable to differentiate between unreliable iPeers and reliable ones and thus must take any beliefs based on those intuitions to be defeated.

**Intuitions**: Our gut-reactions or initial feelings about a claim.

**General moral intuitions**: the set of all moral intuitions; I argue that they are largely unreliable.

**Near-consensus intuition**: An intuition that is widely shared by most intuiters. For an intuition to be a near-consensus intuition, a certain percentage of people must intuit that it is true. What exactly that percentage is is a question for epistemologists, but it must be well above 50%, and does not have to be 100%. If I had to guess, I’d say it’s about 85%.

**Judgment**: A belief about a claim that is arrived at by means of our rationality; a belief that involves our ability to reason logically.

**Rationality**: Our ability to reason logically. The process of rationality is what is involved in moving from premises to conclusions in arguments; it is our ability to recognize whether claims follow from other claims, or are more likely to be true in light of some evidence, and so on.

**A-intuition**: An intuition for which there is a credible underlying evolutionary explanation.

**B-intuition**: An intuition for which there is no credible underlying evolutionary explanation.

**R-intuition**: A claim that is justified solely by appeal to our intuitions. I would argue that it is an R-intuition that morality exists, which is to say, *prima facie*, that we should-E believe that there is such a thing as right and wrong actions. A claim is an R-intuition iff it is a near-consensus B-intuition.
**Rational judgment:** A claim that is ultimately justified by appeal to both R-intuitions and the process of rationality. A claim can be demonstrated to be a rational judgment in multiple ways: one, it can be shown to logically follow from an R-intuition; two, its negation can be shown to conflict with an R-intuition; three, it can be shown to directly entail from the truth of another rational judgment (A) and rational judgment A follows from the truth of an R-intuition; four, its negation can be shown to conflict with another rational judgment (A) and rational judgment A follows from the truth of an R-intuition; five, it can be shown to directly entail from the truth of another rational judgment (A) and rational judgment A follows directly from the truth of another rational judgment (B) and rational judgment B follows from the truth of an R-intuition; and so on…

**R-idea:** A claim that is either an R-intuition or a rational judgment. It is sometimes unimportant to identify precisely how a claim gets its justification, as the argument in question works regardless of whether the claim is an R-intuition or a rational judgment, so I will refer to some claims as being simply ‘R-ideas.’

**Hidden Intuition:** An intuition that is reliable but isn’t an R-intuition. It is “hidden” in the sense that such intuitions are likely to exist but will be overlooked by my method of reflective equilibrium.
Contents:

1. On the Relevancy of our Moral Intuitions........................................ 5
2. On the Correct Method for Seeking Moral Truth.............................. 26
3. Reflective Equilibrium in Practice: Deontology? ............................. 31
4. Or Consequentialism? ................................................................. 47
ABSTRACT

In this project, I seek to an answer to the question of whether the correct ethical framework is a deontological or consequentialist one. I begin with the assumption that morality is objective. Using that as a starting point, I discuss the two primary sources of data on moral topics: our intuitions and our judgments. I note some epistemic worries with regards to our moral intuitions, and argue that the process of arriving at judgments—rationality—constitutes a reliable process. Then, I argue for a method of doing ethics known as “reflective equilibrium,” which I modify in light of some concerns about the reliability of our intuitions and judgments. Next, I apply the method of reflective equilibrium to the deontology/consequentialism debate, and conclude that consequentialism is the correct ethical framework.

I. ON THE RELEVANCY OF OUR MORAL INTUITIONS

Before proceeding, I would like to make an important distinction between the notion of an intuition and that of a judgment. By ‘intuition’ I am referring to the initial compulsions we have regarding possible answers to various questions that are in some sense prior to argument and reasoning. For instance, consider the following scenario, presented by Peter Singer in his article “Ethics and Intuitions.”

Mark and Julie are brother and sister and on a summer vacation in France. One night, they decide that it would be interesting to try and make love. They use two forms of birth control, they both enjoy the sex and think that it has brought them closer together, but they decide never to tell anyone about it and also never do it again. The reaction of disgust that is typically engendered by this scenario captures what I am referring to by ‘moral intuition.’ The average person, confronted with this scenario, would immediately react by condemning the actions of Mark and Julie as wrong; indeed, I did so myself when I first read it. Our intuitions are somewhat akin to our gut reactions to claims or thought
experiments—they are our initial inclinations to think or believe something about an idea that occur separately from our rationalizations about that idea.

Our rationalizations, on the other hand, refer to what I have termed our ‘judgments.’ Judgments are beliefs that we form by means of our rationality, and can easily diverge from our intuitions. Imagine for a moment that when you initially examine the scenario with Mark and Julie, you have the moral intuition that the sex act was wrong for them to perform. However, after thinking about the scenario, you recognize that their actions have almost assuredly caused net positive utility (they have a stronger bond as siblings; Julie will not get pregnant and produce a genetically malformed child; they both enjoyed themselves; there is no risk of a future performance of the act; nobody else will hear about it and react with disgust) and then conclude that their action was morally right. That conclusion is what I term a ‘judgment’—it is an idea that is produced by means of reasoning and thought, and does not arise on a purely instinctive level.

While in the above scenario the judgment and intuition I mentioned involved different answers to the question of whether Mark and Julie did something wrong, our judgments and intuitions do not always diverge, and different people do not always have the same intuition or judgment on a given topic. Some people will certainly have the judgment that what Mark and Julie did was wrong; for example, they might think, “What they did makes me feel disgusted, so I clearly have the intuition that what they did was wrong. Since trusting my intuitions constitutes a reliable process for producing true moral beliefs, I conclude that what they did was indeed wrong.” On the other hand, although likely rare, it is possible that someone will have the intuition that what Mark and Julie did was right.
In this section, I will argue that, generally speaking, our moral intuitions are not reliable indicators of moral truth and thus should-E\textsuperscript{a} be given little evidentiary weight, and then I will argue that the process of reasoning, and using our rationality, is a reliable process for producing true moral beliefs. However, this does not imply that all of our moral intuitions should-E be ignored when evaluating moral theories—I will argue in the following section that there is a small subset of moral intuitions that we should-E take to be reliable. Additionally, that our rationality constitutes a reliable process does not imply that all our judgments should-E be involved in our evaluations of moral theories; even if the process itself is reliable, it can still be led astray if it is applied to unreliable premises. Later on in this project I will delineate the conditions required for a judgment to be considered reliable under my methodology.

There are a few different reasons for thinking that moral intuitions are unreliable. One involves a problem similar to that of peer disagreement in epistemology. In polling a variety of people, we would expect a good amount of variation in their intuitive reactions to various moral questions or scenarios. For such variant intuitions to be epistemically fruitful (i.e. for them to give us evidence for thinking that some related moral claim is true), we need a way of eliminating a good amount of the divergence between intuitors by labeling it as coming from an unreliable source, and for the remaining intuitions to widely converge on the answer to some candidate moral question. For instance, assuming for a moment that moral intuitions are generally reliable, it is clearly unhelpful to a seeker of ethical truth to discover that, on the question of whether action X is right to perform, half of the general population intuits ‘no’ and

\textsuperscript{a} Should-E means ‘should’ in the epistemic sense. The epistemic norm tells us to seek true beliefs and avoid false ones; thus, if one ‘should-E’ do X, then doing X would help one acquire true beliefs and avoid false ones.
the other half intuits ‘yes.’ In order for the seeker to gain good evidence in favor of an answer to the question of whether action X is right to perform, he needs to narrow the set of intuiting individuals to a set that is both more reliable and more convergent around a specific answer. To successfully do so, he would need to find a good way of distinguishing between the intuitive reliability of sets of individuals and also hope that, after eliminating any data from those found to be “less reliable,” there is a sufficient level of convergence around a particular answer for him to be justified in believing that answer. What such a level of convergence might be is certainly above 50% (we should-E do better than chance in deciding what to believe), but it might also need to be much higher (although that is a question for another paper).

It is very difficult to distinguish between the reliability of individuals when it comes to their moral intuitions, because it is not apparent that things like intelligence or level of information have a clear role to play in making such intuitions more likely to be true. Moral intuitions are akin to our pre-rational gut reactions to scenarios or moral ideas, leaving little role for intelligence to play; furthermore, while I have noticed that my judgments on moral questions have changed as I have gained in my total information on moral subjects, the same cannot be said of my moral intuitions: those have remained relatively stable, I just have afforded them less weight. If this phenomenon applies to many others who study ethics, which I expect it might, it gives us further reason for thinking that our level of information on moral topics also has little bearing on the general reliability of our moral intuitions. So our seeker of moral truth needs to discover some other distinctions that have bearing on the relative reliability of an individual’s moral intuitions, or else he should-E accept this problem of “iPeer disagreement” involving

\[\text{\textsuperscript{b}}\text{I do not want to label this problem just “peer disagreement,” because ‘peer’ is an epistemic term that refers to people with roughly equal intelligence, attentiveness, and level of information on a given topic, whereas I have just argued that intelligence and information levels do not have}\]
divergent moral intuitions as a defeater for the justification of any moral beliefs that are based in intuitions that do not have near-consensus support within the general intuiting community. I doubt that our seeker of moral truth can successfully differentiate between the reliability of the intuitions of enough iPeers\(^c\) to successfully surmount this problem. As such, even if there are some moral intuitions that stem from reliable sources, we have a (potentially insurmountable) epistemic problem in that we are unable to adequately distinguish the reliable intuitions from the unreliable ones. It follows that all but “near-consensus intuitions” (i.e. intuitions that are shared by a significant percentage of iPeers) should not be used as evidence in favor of a related moral proposition.

A defender of the general reliability of moral intuitions might reply to my worries in the following way: perhaps in cases where our moral intuitions diverge significantly, we cannot trust one set of intuitive reactions over another and thus we find any moral beliefs on such cases defeated (assuming that they rest substantially on intuitive reactions for their justification). However, not all cases involve divergent intuitions, and there might even be a large number of moral questions for which there is a preponderance of convergent intuitions as to what the right answer to the question is. For example, most iPeers have intuitions that comport with the proposition that murder is wrong. So in all such near-consensus cases, we are justified in believing that the near-consensus answer is the correct one solely because the vast majority of iPeers intuit that it is so.

In light of this objection, a challenge still remains for the defender of the reliability of moral intuitions. This challenge involves a historically popular set of thought experiments that much bearing on the reliability of an individual’s intuitions. So I will refer to “peers” in terms of intuition-reliability as “intuiting-peers,” abbreviated using the more hip locution “iPeer,” and will term to the epistemic problem discussed here “iPeer disagreement.”

\(^c\) Two people are iPeers iff they are both capable of having moral intuitions.
philosophers have referred to as the “trolley problem,” which I will argue demonstrates that our near-consensus intuitions are prone to inconsistency and thus, *prima facie*, should-E be taken to be unreliable.

Phillippa Foot presented the first version of the trolley problem in a 1976 article about the Doctrine of Double Effect. The Doctrine of Double Effect is used to determine the moral status of actions that involve both good and bad effects. Foot motivates belief in the Doctrine by claiming that there are two cases that can be properly distinguished only by appeal to the Doctrine, which essentially holds that it is always wrong to do a bad act intentionally in order to bring about good consequences, but it is sometimes permissible to do a good act despite knowing that it will bring about bad consequences. The second case that Foot presents as part of her argument has spurred a lot of contemporary moral thought, and it runs as follows:

A runaway trolley is barreling down a track, and at a fork there are 5 people working on one side and 1 person working on the other. The trolley is headed for the side with five people, but a lever can be pulled to switch the trolley’s path. Either path chosen by the controller will result in the death of all the workers on that side of the track. The controller of the trolley can decide whether or not to switch the track, killing one to save 5.

Foot thinks that the Doctrine of Double Effect successfully interprets this case, which I will call Trolley, by providing justification for the most common intuitive reaction that people have when presented with the case: that the controller should-M\(^d\) switch the track, killing one to save five. I will assume that this intuition is a near-consensus intuition (i.e. a sufficiently large percentage of people have this intuition) for the purpose of this argument. Bolstering my assumption here, a 2006 survey of some 60,000 subjects spanning 120 countries indicated that about 90% of people intuit that it is permissible for the controller to switch the track in Trolley.\(^3\)

\(^d\) ‘Should’ in the moral sense. The moral norm tells us to do things that are morally right and to not do things that are morally wrong. Thus, if one ‘should-M’ do X, that implies that it would be morally right for one to do X; this is the same as saying “one ought to X.”
The case that has often been paired with Trolley is called Footbridge. Footbridge runs something like this:

A runaway trolley is barreling down the track, and further down the track there are 5 people working. You are standing on a footbridge overlooking the track, and in front of you is a very large man who is bending over the railing and watching the trolley advance. There are no other people around, and you have the option of pushing the fat man into the path of the trolley, killing 1 to save 5 (he is large enough to disrupt its momentum completely), or doing nothing, allowing the 5 workers to die.

The most prevalent intuitive reaction to Footbridge is that you should-M not push the fat man into the path of the trolley, and I take Footbridge, like Trolley, to be a near-consensus case. Here, data also indicates that about 90% of people have the intuition that you should-M not push the fat man, so this again seems like a good assumption. These same findings (that 9/10 people have what I’ve termed the ‘near-consensus’ intuitions in Trolley and Footbridge) were also replicated in a separate poll presented in an episode of RadioLab.

Trolley and Footbridge present a challenge for the defender of the reliability of moral intuitions. Since they constitute near-consensus cases, the defender wants to say that both of the predominant intuitions are correct—it is right to switch to the track with only one worker in Trolley, and wrong to push the fat man in Footbridge. The challenge is for our defender to identify a morally relevant difference between the two cases that can justify the claim that it is morally right to kill one to save five in Trolley and morally wrong to do so in Footbridge. I will argue that no such difference between the cases is present, and consequently our near-consensus intuitions are sometimes inconsistent and thus, ceteris paribus, near-consensus intuitions should-E not be construed as reliable.

To motivate my claims here, I will explore some candidate morally relevant differences between Trolley and Footbridge and explain why I do not find them compelling. The first is that
in Trolley the controller is not physically touching the person that he kills, but in Footbridge you physically shove the fat man to his death. However, imagine two scenarios in which a moral agent, let’s call him Dwight, is murdering Jim. Both scenarios are identical except for the fact that in scenario 1, Dwight kills Jim by strangling him, and in scenario 2 he kills Jim by shooting him in the head with a shotgun. If physical touching constitutes a morally relevant difference, it seems as though Dwight’s action in scenario 1 is morally worse than Dwight’s action in scenario 2, but this is clearly not the case. Hence, it is entirely unintuitive to think that physical touching constitutes a morally relevant difference between the cases.

Before proceeding, I would like to note a potential objection to the reasoning I just employed in the previous paragraph. One might claim that if I am arguing against the reliability of moral intuitions, I cannot justifiably take our intuitions to be good evidence for the proposition that Dwight’s action in scenario 1 is not morally worse than Dwight’s action in scenario 2, as it seems as though I have just done. However, I think that this objection is unproblematic because either the objection does defeat my argument here and hence our near-consensus intuitions are unreliable (as that is an assumption involved in this objection), or else it does not apply and I do have good evidence for claiming that Dwight’s action in scenario 1 is not worse than his action in scenario 2, and my argument can proceed unhindered.

A second proposal for a morally relevant difference between Trolley and Footbridge is that all six workers in Trolley are to some degree “in harm’s way,” whereas the fat man cannot be so construed. It is important to note that without any intervention by the controller in Trolley the solitary worker would not be harmed; however, there might be something to the idea that the worker in Trolley is in a situation where it is more likely that he will be killed in a trolley
accident than the fat man in Footbridge, and as such he is more “in harm’s way” than the fat man.

This “in harm’s way” proposal also seems wrong. Imagine Dwight again murdering Jim, this time by poisoning a greasy cheeseburger that Jim is eating. In scenario 1, Jim has really high cholesterol and as such is at an increased risk of heart attack (especially when he eats greasy foods like his daily cheeseburger). Dwight successfully poisons Jim’s cheeseburger and kills him. In scenario 2, Jim does not have high cholesterol and thus is not as likely to die just from eating a cheeseburger. Jim bites into the greasy treat and subsequently dies—Dwight successfully poisoned it and killed him. Does Dwight’s action in scenario 1 seem less bad than his action in scenario 2? Clearly not: he murders Jim in precisely the same way in both scenarios. But in scenario 1, Jim was more likely to be killed from eating a cheeseburger than he was in scenario 2, antecedent to any action by Dwight. If our intuitions here are reliable (and if they are not, then my ultimate claim is proven nonetheless) then being “in harm’s way” cannot count as a morally relevant difference between the Trolley and Footbridge cases.

A third proposal for a morally relevant difference between Trolley and Footbridge involves the Doctrine of Double Effect itself. In Trolley, the controller is not intending to kill the solitary worker as a means for achieving the desired result (saving the lives of the five workers), but rather that event is a foreseen but undesired and unnecessary consequence of achieving that result. (It is unnecessary in the sense that all that is required for saving the five is a diversion of the trolley; if the one somehow avoided his death by, say, getting off the track before the diverted trolley hits him, his miraculous survival would not subsequently cause the five to die). In Footbridge, on the other hand, your pushing the fat man constitutes using his life as a means for achieving the desired result of saving the five other workers. While one might
claim that his death is also unnecessary and undesired for achieving that result—after all, you would be perfectly content if his body successfully disrupts the motion of the trolley without killing him—a philosopher like Foot would note that pushing him into the path of a runaway trolley is “close enough” to killing him that we can reasonably understand your action to involve using his life as a means for achieving the desired result. (Additively, his death is “necessary” for realizing the good outcome because if the fat man somehow avoided his collision with the trolley then the five would die.) Consequently, we can understand your action in Footbridge as “aiming” at the death of an innocent person, which is distinct from our understanding of the controller’s action in Trolley.

To determine whether this last proposal can justify our intuitions in Footbridge and Trolley, I would like us to consider another example that has its roots in Foot’s article, which I will call Cave. A fat man is leading a group of five people out of an ever-narrowing cave, and gets stuck, trapping the group behind him inside the cave. Water levels inside the cave are rising, and to avoid drowning the people inside have only one option: blowing up the fat man with a stick of dynamite. Now imagine two scenarios. In scenario 1, the fat man’s head points out towards the exit, and as such he will not drown with the rest of the group and will eventually get out alive if he is not blown up with the dynamite. In scenario 2, the fat man’s head points into the cave and so he will surely end up dead: either he will drown with the rest of the group or they will blow him up and escape. Now, regardless of one’s intuitive reaction as to what the group should-M do in scenario 1, it seems clear that most people would find that they should-M blow up the fat man in scenario 2…after all, the fat man is going to die anyway, so the rest of the group might as well survive the ordeal instead of perishing with him. However, the Doctrine of Double Effect would note that in both scenarios blowing up the fat man involves using his life as
a means for achieving the desired effect (the survival of the group) and involves the intention of killing the fat man. The Doctrine, if it is to be used to distinguish between the cases of Footbridge and Trolley, needs to (at the very least) hold that it is wrong to intentionally kill an innocent even if one would save the lives of five others by doing so, and as such it would counsel the group in Cave to refrain from blowing up the fat man in both scenarios 1 and 2. Since the predominant intuitive reaction to scenario 2 would be that the group should-M blow up the fat man, I find that the Doctrine of Double Effect is inconsistent with our intuitions, and thus either our intuitions are unreliable or else the Doctrine of Double Effect is false (and my ultimate claim follows either way).

One can also argue that the Doctrine of Double Effect is unrelated to the factors that actually lead to the differing near-consensus intuitions between Trolley and Footbridge. Take for example ‘Loop,’ another version of the Trolley Problem that runs as follows:

A runaway trolley is barreling down a track towards 5 workers that are a ways down the line. Halfway between the trolley and the workers is a sidetrack that contains a loop. The track can be switched to the sidetrack such that the trolley goes around the loop before returning to its original track. The controller of the runaway trolley notices that there is a fat man in the middle of the loop. If the controller switches the track such that the trolley goes around the loop, it will hit the fat man and halt, leaving the 5 workers unharmed; if he does nothing then the train will miss the loop and the 5 workers will die.

The near-consensus intuition in Loop is that the controller should-M switch the track, using the fat man’s life as a means for saving the five workers. Given that result, it can be said that most people find little moral difference between the cases of Loop and Trolley, even though the fat man’s death is necessary for achieving the good result in Loop, whereas the death of the one is not necessary for achieving the good result in Trolley. This finding would indicate that, in Footbridge, the use of the fat man’s life as a means for achieving the desired result is not what
leads most people to intuitively condemn pushing him in front of the trolley. It is more likely that the differences of physical touching or being in harm’s way are what led to that intuition—but as we have already seen, those differences are not morally relevant. Thus, Loop can be understood as another thought experiment that undermines the intuitive support for the Doctrine of Double Effect.

Three proposals for distinguishing between Trolley and Footbridge—physical touching, being in “harm’s way,” and using an innocent person’s life as a means for achieving an end (i.e. intending to kill an innocent person)—have all failed to justify our seemingly inconsistent intuitions between the cases. However, these three proposals do not constitute an exhaustive list of all the possible morally relevant differences between Trolley and Footbridge. Other proposals might point to differences in physical distance, notions of responsibility, or something else entirely. However, I am not seeking to definitively answer the classic “trolley problem” here, but merely to demonstrate that the onus is on the defender of the reliability of our moral intuitions to solve the dilemma—because as far as I can tell, there is no morally relevant difference between the cases which can serve to justify the near-consensus intuitions that the controller’s action in Trolley should-M be to switch to the track with one worker and that your action in Footbridge should-M be to refrain from pushing the fat man.

I would also note that any future proposal for a morally relevant difference between the cases would not only need to demonstrate why that difference should-E be construed to have moral relevancy, but also (and this is an even more difficult challenge) needs to explain why it is of such great moral relevancy that is can switch the moral status of the “killing-one-to-save-five”
Clearly, the common intuiter would construe a single life to be a very valuable thing, and thus the lives of five people to be even more valuable… so whatever proposal a defender of the reliability of our moral intuitions comes up with would need to involve a difference that has clear and substantial intuitive significance, enough to make it so that the preservation of one life becomes the moral imperative, even when the actor might preserve five lives by taking the one. That no such proposal has been discovered after a fair bit of examination indicates, to me, that such a difference is simply nonexistent, and so I conclude that Trolley and Footbridge demonstrate that some of our near-consensus intuitions are inconsistent. The finding that even near-consensus intuitions can be unreliable counts as a further strike against attaching evidentiary weight to our moral intuitions.

Thus far, I have argued that our moral intuitions are unreliable, in part because of the divergence of moral intuitions in many cases and in part because they can lead us astray even in near-consensus cases. The other side of this argument, and the final nail in their coffin, is to provide an explanation for why moral intuitions are unreliable. If we are able to provide a compelling explanation for why they are unreliable, then I think that we have sufficient evidence for claiming that we should not grant our moral intuitions much evidentiary weight. For this explanation, I turn to an argument presented by Peter Singer in his article “Ethics and Intuitions.”

---

e To illustrate my point here, consider two scenarios in which Dwight murders 100 people. In the first, he does so just for fun. In the second, he does so in order to save the life of his daughter. While his difference in purpose is (intuitively) morally relevant, it is also clearly insufficient for switching the moral status of Dwight’s action: he committed moral wrong in both scenarios.
Singer argues that our moral intuitions are, in a sense, the byproducts of evolution. If Singer is correct, then such a finding would demonstrate that our moral intuitions are unreliable. This is because evolution tracks what I call the “gene survival norm,” and the gene survival norm and the moral norm are not coextensive. If we assume that our moral intuitions are evolutionarily engendered, then the sorts of intuitions that we would tend to develop over time would be ones that facilitate our survival and reproduction in whatever cultural environment we are living in—in that sense they follow the gene survival norm. Furthermore, any ideas or actions that tend to help us (individually) survive and reproduce are not necessarily identical to the ideas or actions that are morally right for individuals to have or do. I can think of many cases in which it is obviously wrong for someone to perform an action even though that action is clearly in the interest in terms of their survival or reproduction—in fact, evolution shaping our moral intuitions might even be responsible for a large amount of the human vice of selfishness.

As an extreme example, imagine that Michael has a gun to his head and is facing a button that, if pressed, would launch a nuclear warhead that will blow up the city of Manhattan. If Michael does not push the button, then the gunman will kill him. Thus, Michael can choose to persist by pressing the button and launching the missile that will blow up Manhattan (which, assuming that Michael has no blood relatives there, is what the gene survival norm would tell him to do), or he can choose to be shot with the understanding that Manhattan will not be blown up. Even though it is not in his interest in terms of genetic reproduction, the better thing for Michael to do is—clearly—to allow himself to be killed. Cases like these motivate the claim that the gene survival norm and the moral norm are not coextensive.

So what reasons do I have for thinking that our moral intuitions have their origin in evolution? Singer’s argument, which I support, runs something like this. First off, it seems to
explain the shape of many of our intuitions. Take, for instance, the fact that the average person loves his own children more than his cousins, and his cousins more than strangers. Singer claims, and reasonably so, that genes that lead to these forms of love are more likely to survive and spread among social mammals than genes that do not. This same idea would also explain why we sometimes experience revulsion when hearing about people who neglect their own children to the point of death, but tend not to feel such revulsion towards people that fail to donate money that could prevent the starvation of children in third world countries. Generalizing from these explanations, I think that we can come up with a compelling evolutionary story behind many of our more commonly shared moral intuitions, which lends credence to Singer’s claim—that our moral intuitions do indeed have their origin in our evolutionary history.

The second part of Singer’s argument is to note that we share some moral compulsions with other species, like monkeys and chimpanzees. As one example among many, take the moral intuition that we should repay favors. Singer describes studies done on monkeys that have found that when a monkey picks out another’s parasites, if the other fails to return the service then they are often attacked. The prevalence of seemingly moralistic behavior in animals presents us with a possible connection between human features of morality and features of the social structures of animals. Hence, the seeming existence of basic moral intuitions or compulsions in other less advanced biological species, and the fact that those compulsions clearly facilitate the survival of those species, serves as another piece of evidence in favor of an evolutionary explanation for the origin of our moral intuitions.

The final, most convincing piece of evidence presented by Singer for the influence of evolution on our moral intuitions involves his discussion of the results of fMRI experiments conducted at Princeton by Joshua Greene. These experiments involved brain scans of people
pondering over the dilemmas presented in the Trolley and Footbridge thought experiments. The results of the imaging studies confirmed what Greene had hypothesized at the beginning of his research—the people who were asked to make a moral judgment in the Footbridge case showed more activity in parts of the brain associated with emotion, and those who were asked about Trolley had more activity in parts of the brain associated with logical reasoning. Furthermore, Greene found that the few who considered it right to push the fat man in Footbridge took significantly longer to reach this judgment than those who answered that it would be wrong to do so, presumably because they were forced to reason beyond their primitive, biologically driven moral intuitions. These sorts of results demonstrate the connection between our brain and our moral intuitions in a way that would indicate that our moral intuitions are evolutionarily engendered.

The best explanation for all these pieces of evidence—the general comportment of our moral intuitions with explanations about what would have been in our evolutionary interest throughout our ancestral past, the existence of similar moral compulsions in other species, and neurological findings which tie intuitive judgments to the structure of the brain which is involved in processing emotion—is that evolution has played a significant role in the origin and development of our moral intuitions. However, this explanation also indicates that our moral intuitions are “tainted” by their connection with the gene survival norm—and, as such, cannot be taken as reliable indicators of moral truth.

The epistemic consequences of the unreliability of our moral intuitions will be more fully discussed later on, but it certainly has a large influence on how we should seek moral truth. It seems as though we need to move beyond our moral intuitions if we are to come to a firmer understanding of morality.
The second claim that I am going to argue for in this section is that our rationality—our ability to reason logically—is an epistemically reliable process. Recall my earlier characterization of judgments: that they are the products of rationality, and can (but do not need to) diverge from our moral intuitions. Unlike moral intuitions, they are going to involve beliefs for which people can more easily provide explanations. While I would not claim that, prima facie, the content of any judgment is itself reliable (since it is not just a product of rationality but also is dependent on the propositions from which it is “reasoned,” and thus also depends on the reliability of those propositions), I am claiming here that the process of rationality is itself a reliable process. This would entail, for instance, that if we arrive at judgments by means of reasoning from certain “foundational propositions” about morality that we take to be intuitively true, those judgments are not reliable since our intuitions are not reliable indicators of moral truth, but if we are able to discover foundational propositions about morality that we should-E believe to be true then we should-E believe the judgments that follow from those propositions.

Before proceeding, I want to consider an important objection to my claim that our rationality constitutes an epistemically reliable process. All of the following claims are likely true: we can come up with a good evolutionary explanation for the development of our ability to reason, certain species of animals exhibit behavior that indicates that they have some reasoning ability, and there is neurological evidence that our reasoning abilities are connected to our brain development. If these kinds of claims undermine the reliability of moral intuitions then why are they not concerning for the reliability of our process of rationality?

The reply is that it is not epistemically worrisome for the process of rationality to track the gene survival norm, unlike our moral intuitions. For our reasoning capacity to be evolutionarily advantageous, and thus for it to develop to the extent that it has throughout the
development of our species, it needs to reliably get us to knowledge about the way the world actually is. Imagine an ancient tribe of humans at war with another tribe. In order for the members of one tribe to survive and pass on their genes, they need to emerge victorious. In terms of achieving this goal, it certainly helps a tribe if they have the ability to out-strategize and outmaneuver the other tribe… but that is only possible if the tribe members have reasoning capacities that can reliably get them to judgments that are correct about the way the world actually is (for instance, where the other tribe is likely to plot an ambush, or position their warriors, or what weapons they will be using). The correspondence theory of truth understands truth to be determined by the way the world actually is—and the way the world actually is is exactly what our process of rationality needs to be good at getting us to in order for it to be evolutionarily advantageous.

My argument here can be boiled down thusly: either our ability to reason is evolutionarily engendered or it is not. If it is not, then the epistemic worry about moral intuitions does not apply to our rationality, and thus it can be fairly construed as reliable. If, on the other hand, it is evolutionarily engendered, then our rationality is an evolutionarily advantageous trait. If our rationality is an evolutionarily advantageous trait, it is because it does a good job of helping us arrive at propositions that correspond to how the world actually is. But the correspondence theory of truth defines truth as propositions that comport with how the world actually is. So if our rationality is evolutionarily advantageous, then it must do a good job of helping us get to truth. If our rationality does a good job of helping us get to truth, then it is a reliable process. So regardless of its origin, it follows that our rationality indeed constitutes a reliable process.

I would also add that this argument does not apply to moral intuitions because the claim ‘if our moral intuitions are evolutionarily advantageous then it is because they do a good job of
helping us arrive at propositions that correspond to how the world actually is’ is false. Our moral intuitions are advantageous because they led our ancestors to make judgments that facilitated communitarian living regardless of whether those judgments were objectively true or not. For instance, it helps a community survive and procreate if its members have beliefs that lead them to favor their own community members over members of rival communities and if they think that their kids and neighbors deserve special moral consideration over strangers and foreigners, but such beliefs fly in the face of candidate objective moral truths like “all human beings have equal moral weight.” However, it does not help those communities survive if their reasoning capabilities consistently fail to lead to beliefs in propositions which correspond with the way the world actually is, which is why the reliability of the process of rationality is preserved even if it is the byproduct of evolution.

One might object that it would theoretically help communities survive if our reasoning capabilities strengthened our beliefs in the sorts of moral propositions that are supported by our moral intuitions; hence, it also seems as though our rationality could be evolutionarily advantageous without being good at getting us to truth. While this objection is certainly plausible, it ignores the fact that in most domains of life our rationality would be disadvantageous if it did not engender judgments that comport with the way that the world actually is, which means that if it were not on the whole “truth-conducive” then overall rationality would not be evolutionarily advantageous and thus would likely not have developed. For instance, if we historically were unable to reliably reason about what sorts of tracks are fresh and indicate the presence of nearby prey, we would have been poorer hunters—as often chasing after animals that have long since disappeared as those that were actually nearby. If we were unable to reason about how our companions were using language, we would have had a much
harder time coordinating actions with other humans during group hunts, in battle, or while setting up social institutions that facilitated community survival. The potential examples here are endless—clearly, the evolutionary advantage underlying our rationality lies in it being efficacious in getting us to truth.

A much more interesting objection is that our rationality has led us to all sorts of different and even contradictory claims throughout human history. One merely needs to survey a sliver of the field of philosophy to find a wide range of disagreement about the truth of various judgments. I think that this objection essentially boils down to the epistemic problem of peer disagreement: the idea that a judgment is defeated whenever someone with an equal level of information, attentiveness, and intelligence on a given topic disagrees with that judgment. Since this phenomenon occurs frequently (and across a host of topics), it seems as though most philosophical beliefs will lack justification. However, I doubt that this problem defeats my claim that rationality constitutes a reliable process. I think that a lot of the variance in the judgments that issue from the use of rationality can be explained by appeal to things like informational inputs from unreliable sources (i.e. our moral intuitions), variations in information levels, and differences in intelligence; thus, we do not have to construe this variation as an indication that our process of rationality is itself unreliable—not to mention the fact that peer disagreement may not be as big of a concern in the future once more knowledge has been accumulated. For example, I have argued that our moral intuitions are unreliable, which would explain some of the variation in the judgments made by various moral philosophers (as instantiations of cases in which the philosophers applied their reason to information which was itself derived from an unreliable source, our intuitions). Furthermore, Singer himself notes in “Ethics and Intuitions” that many of the previous great moral philosophers did not have access to technology that made
things like neural imaging possible; advancements in technology have assuredly (if not steadily) increased our information levels throughout history, meaning that the information that we are able to apply our reason to has both grown in some areas and gotten more precise in others. Since more reliable information would make our judgments more likely to be true, we would expect there to be a measure of variation between judgments made by philosophers that not only have different levels of information than their contemporaries, but also for there to be variation between philosophers over time. Given these separate influences on the content of our judgments, it seems as though one can coherently absorb this objection and still claim that our rationality does constitute a reliable process. After all, if one were under the influence of Descartes’ Evil Demon, the fact that the content of most of one’s judgments is likely to be false is not good evidence for thinking that the use of our rationality is itself unreliable: the trouble clearly comes from the source of the information that we apply our reason to.

Hence, I submit that our rationality constitutes a reliable process. Furthermore, if we take care to examine our sources of information carefully, and select only foundational beliefs that we should-E believe (which we can then build on by means of our rationality, much like how mathematics takes foundational claims and uses them to prove other related claims), I think that we can construe the content of our judgments to be reliable as well—in part because this is the best we can do in the search for truth. After all, we do not have access to the information that will come to light in the future, so the best we can do is use whatever we have now. And if “the best we can do” is insufficient for claiming that our beliefs are justified, then the very project of ethical philosophy is doomed, at least at this point in time. But since the very project of ethical philosophy is not doomed at this point in time, it follows that the content of our judgments can

\[\text{footnote}[f]{I’m just going to stipulate this, because the alternative is abhorrent.}\]
be taken to be reliable so long as we satisfy ourselves as to the reliability of certain foundational beliefs.

II. ON THE CORRECT METHOD FOR SEEKING MORAL TRUTH

I have argued that our moral intuitions are largely unreliable, and that the content of judgments can be construed to be reliable so long as the judgments are grounded in reliable foundational beliefs. Next, I will argue that the methodology we should apply for seeking moral truth is the reflective equilibrium approach, as it is explicated by Sandberg and Juth in their article “Ethics and Intuitions: A Reply to Singer.”

The other half of Singer’s “Ethics and Intuitions,” after he argues (as I have) that our moral intuitions are unreliable, is a rejection of the reflective equilibrium method of doing ethics. Singer characterizes this method as a search for the theory that best fits both our “considered moral judgments” and our “prior moral judgments” (i.e. what I have termed our “judgments” and our “moral intuitions”). However, he thinks that any method of doing ethics that judges a normative theory in part by the extent to which it matches our moral intuitions is fundamentally flawed—as we do not have good reasons for thinking that our moral intuitions are reliable in the first place; as such, there is no point in trying to make sure that our moral theories fit with our moral intuitions whatsoever.

At the end of his essay, Singer notes a difficulty with his argument: it seems as though without intuitions, we can go nowhere. This seems right. A belief in consequentialism must rest, fundamentally, on our intuitions—but how can someone like Singer be a consequentialist and yet reject our moral intuitions as unreliable? Singer notes that we might try to distinguish our immediate emotional responses and our more reasoned conclusions, which is what I have
done in drawing the distinction between our moral intuitions and our judgments. Singer sees this distinction as the “only way to avoid moral skepticism.”

However, I am not inclined to agree: if all moral theories must rest on some sort of fundamental intuition, then we need to make a sort of foundationalist claim about the process of doing ethics if we are to avoid moral skepticism—that there are some kinds of moral intuitions that are justified even though, generally speaking, moral intuitions are unreliable. Judgments can only get us so far, as even if the process of reasoning is itself reliable, the process needs to get its initial data from somewhere. And if the initial data is itself unreliable, the content of our ensuing judgments will be no better (remember Descartes’ Evil Demon). Thus, some moral intuitions need to be taken as reliable if one is to develop a moral theory.

Sandberg and Juth mention a similar idea in their article. They note that while Singer is correct in that basing ethics solely on moral intuitions is problematic, basing it solely on ‘reason’ gives rise to similar problems. As such, the best solution would be to strike a balance between the two, which is what they see as being, fundamentally, the method of reflective equilibrium.

To motivate their claim that basing the search for moral truth solely on reason is problematic, they point to Kant, as he produced the “most systematic normative ethical theory from the basis of pure reason in the history of philosophy,” (2011: 219) and yet his categorical imperative is plagued with endless counterexamples that make it unpalatable as a theory. They also note a commonality between Kant’s normative theory and Singer’s own preference utilitarianism—both are supposed to follow naturally from the universal nature of morality, and yet are widely divergent theories. Thus, either the reasoning employed by Kant or Singer is defective, or their theories are in some way reliant upon (faulty) moral intuitions after all.
I think that, regarding the above disjunctive, it is the latter: both theories are in some way reliant upon moral intuitions. Even if the theories are supposedly derived from pure reason, the reasoning has to begin with some sort of proposition that is simply taken to be intuitively true. Thus, I agree with Sandberg and Juth that our judgments can only get us so far: the development of a moral theory requires that we start somewhere, and that somewhere has to be some sort of moral intuition. For no matter the power of our reason, even the very idea that morality exists stems in part from our intuitions, as it is not an observable phenomenon (even if you witness a brutal murder, you are not going to have a sensory experience of wrongness). However, I do not think that this development necessarily leads us to a form of reflective equilibrium under which all our moral intuitions have to be given some weight when evaluating various normative theories, which is how Singer seems to construe the method, but rather to a form similar to what Sandberg and Juth propose at the end of their article.

Sandberg and Juth look for a possible distinction between moral intuitions that would allow us to claim that some are relatively more reliable than others. Such a distinction would allow us to incorporate only the former into the reflective equilibrium equation and thus, in a sense, “get the best of both worlds”—a methodology that asks moral theories to fit with our

---

Sandberg and Juth also talk about the possibility of a coherentist view of moral justification in which there is no individual statement or principal that must be true, as opposed to the foundationalist account that requires the grounding of moral theories on certain “foundational” intuitions. While the question of whether foundationalism or coherentism about moral justification is correct is outside the scope of this project, I will proceed under a foundationalist approach for practical purposes… a coherentist argument would simply be too monumental a task, as it would require an evaluation of the coherency of consequentialism and a comparison between both it and deontology. However, I think that my argument will have merit under either approach, since if coherentism is true then I can be seen as simply detailing the overall coherency of the consequentialist’s framework; the missing piece is just an evaluation of its level of coherency vis-à-vis deontology’s. Also relevant is Sandberg and Juth’s claim that, under the coherentist’s version of reflective equilibrium, consequentialism has been explicitly defended as the most superior reflective equilibrium.
reliable moral intuitions and also comport with our judgments that are based on those reliable
intuitions. The distinction they settle on is between the contentious cases and the “obvious, non-
problem” cases… such as the intuition that the conductor should-M throw the switch in Trolley.
However, I think that this distinction is going to fail, as the mere fact that an intuition is almost
universally shared and regarded as true gives us insufficient reason for establishing it as
reliable… it still may well be evolutionarily engendered.

The correct distinction between intuitions that we can consider prima facie reliable and
those we cannot is first going to draw a line between moral intuitions that can be explained in
evolutionary terms (which I’ll call A-intuitions) and those that cannot be explained so easily
(which I’ll call B-intuitions). Once that distinction has been drawn, we then want to apply
Sandberg and Juth’s distinction to separate B-intuitions that are contentious from B-intuitions
that are widely intuited to be true (so as to avoid the problem of iPeer disagreement). The B-
intuitions that are widely regarded as true can then be justifiably construed as prima facie
reliable moral intuitions, since they are both “near-consensus” and free from the worry about
evolutionary engenderment. I’ll call these kinds of intuitions R-intuitions. It follows that any
claim is an R-intuition iff it appropriately relies on our intuitions for its justification, which is to
say that it is a near-consensus B-intuition that the claim is true. Any moral intuition that is not an
R-intuition cannot be construed as reliable, given both the epistemic concerns discussed earlier
regarding moral intuitions that are evolutionarily engendered and the epistemic concerns
involving iPeer disagreement. With the distinction between R-intuitions and all other moral
intuitions in hand, we can now construct a reflective equilibrium methodology that can be used
to discover moral truth that circumvents both Singer’s earlier worries about the method and
Sandberg and Juth’s concerns about a methodology that relies solely on judgments.
The methodology that I will use throughout the final sections of this project is as follows: we search for R-intuitions, apply our rationality to them, and discover what judgments ensue from the various R-intuitions. Whatever judgments follow from R-intuitions I will term *rational judgments*, and we can justifiably consider rational judgments to be reliable, just like R-intuitions. Consequently, we will evaluate various moral theories based on their comportment with these rational judgments and R-intuitions. The ethical framework that we should believe to be true is that which fits all rational judgments and R-intuitions the best, relative to other frameworks. And as I will eventually argue, that framework is consequentialism (at least as compared to deontology). This methodology is what I will heretofore understand as the method of reflective equilibrium.

Later on in this project, we will encounter claims for which it is unclear whether they are R-intuitions or rational judgments grounded in R-intuitions, since it is ambiguous whether they rely on our intuitions for their justification or if they rely on our faculty of rationality as applied to R-intuitions for their justification. (Remember, R-intuitions are *prima facie* justified because they are near-consensus B-intuitions; rational judgments are *prima facie* justified because they are judgments that ensue from a reliable process [rationality] as it is applied to R-intuitions.) However, since the method of reflective equilibrium demands that a moral theory comport with both R-intuitions and rational judgments grounded in R-intuitions, it will not always matter that we draw this distinction explicitly. So at times I will simply refer to some claims as R-ideas, an R-idea being a claim that is either an R-intuition or a rational judgment.

Sandberg and Juth demonstrated that a methodology that relies purely on our judgments is flawed, and Singer showed us that we should not evaluate moral theories purely on the basis
of their comportment with our moral intuitions. Consequently, if we are to avoid moral skepticism, we need a methodology that can circumvent the concerns of Sandberg and Juth as well as Singer. As I have designed it, our method of reflective equilibrium does exactly that, and since moral skepticism is false, I would also argue that it is the correct method for seeking moral truth (or at least close enough to the correct method that it can be construed as a reliable method for seeking moral truth).

III. REFLECTIVE EQUILIBRIUM IN PRACTICE: DEONTOLOGY?

The question immediately becomes what intuitions, if any, are R-intuitions. The first moral intuition that I propose is an R-intuition is that morality is universal: that whatever moral imperatives exist, they apply equally to all moral agents, and all moral agents are of equal moral importance. I will call this intuition the universal intuition. (I label it an R-intuition because I think it is justified just by appeal to our moral intuitions, which separates it from rational judgments; it is justified because—as I’ll presently argue—it is a near-consensus B-intuition).

The universal intuition is a B-intuition (it cannot be easily explained in evolutionary terms). In fact, it seems to lie in direct conflict with the sort of communitarian living that our ancestors engaged in. In order to survive in most places on Earth, humans would need to favor their own communities far more than the communities of “outsiders,” so that they would be more

---

\(^h\) The correct method would also capture in its net those moral intuitions that are reliable but are not R-intuitions (as it seems likely that there is at least one such intuition). However, since we do not have good distinguishing reasons that can separate these “reliable” intuitions from the rest (for instance, how can we distinguish the moral intuitions that really are evolutionarily engendered from those that are not but still have a plausible evolutionary explanation attached to them?), we cannot, in our current state, justifiably construe any but R-intuitions to be reliable. Thus, the method of reflective equilibrium that I have described here is “the best we can do” with the information that we have, and so in that sense is the correct one, but is not the best method in the theoretical sense.
willing to commit violence in order to capture or defend scarce resources and territory. Additionally, it also helps a human pass on their genes if they tend to attach more moral significance to their own children than to other people. Since the universal intuition clashes with these evolutionary drivers, it is a B-intuition.

The universal intuition is also widely intuited to be true; in fact, it seems clear that the onus is on the objector to explain why some humans do not count as much as others, morally speaking. While one might at first object that serial killers, for instance, do not count as much as other humans, one can still easily explain why we should punish serial killers even if they count as much as other humans by appealing to justifications like incapacitation, deterrence, or rehabilitation. One such justification runs as follows: in the case of the serial killer, removing him from society and depriving him of his freedom might be equivalent to doing something bad to him, but allowing him to roam free would be equivalent to causing more overall harm; thus, we can justify his imprisonment even if we allot him equal moral weight by simply pointing to the notion that his life is not as important as the lives of multiple others. Since the only potential reason for thinking that we should find that serial killers do not have as much moral value as other moral agents is to enable us to justify their punishment, but we can still justify their punishment without construing them as having less moral value as another person, it is not the case that we should find that serial killers do not count as much as others and this objection fails.

Thus, we have discovered our first R-intuition, the universal intuition. Under the method of reflective equilibrium, we construe this intuition to be prima facie reliable, and can demand of candidate moral theories that they comport with it. For example, if a moral theory would tell us that, ceteris paribus, our race, our nation, our neighborhood, or even our family is inherently
more important than any other race, nation, neighborhood, or family, then it should be rejected. This does not necessarily mean that we should not favor our own family over other families, but it does mean that, for example, if there is only one food source left in the world, then all things equal your family does not have any more claim to the food source than another family. A credible moral theory might still claim that there are good reasons for favoring one set of individuals over another set, but it cannot claim that a set of individuals qua a set of individuals is inherently special. Put another way, if theory X says that we should favor set A over set B because set A has special property P* and set B lacks P*, then theory X must also say that we should favor set B over set A if it were the case that set B had the special property P* and set A lacked P*.

Next, I would like to apply the moral knowledge obtained thus far to the deontology-consequentialism debate. As I see it, this debate hinges primarily on whether the moral status of an action is determined by the type of action it is, or by the kinds of consequences it produces. Imagine two scenarios. In scenario A, Jim goes home and says five Hail Mary’s after he promised Dwight that he would not say any (for a trivial reason). In scenario B, Jim goes home and says five Hail Mary’s, but made no such promise to Dwight. There are no other differences between the scenarios, and the way the future unfolds is exactly the same between them. While the typical consequentialist theory would find no morally relevant difference between the two scenarios, the typical deontological theory would condemn Jim’s action of saying five Hail Mary’s in scenario A but not in scenario B, since in scenario A Jim “broke a promise,” which is an action-type (i.e. he performed an action that is an instantiation of promise-breaking).
I think that most people would have the moral intuition that “promise-breaking Jim” in scenario A did something morally worse than non-promise-breaking Jim in scenario B. Indeed, if I had to actualize one of these scenarios, I certainly feel more inclined towards scenario B. However, as we have seen, the presence of a moral intuition is not any indication of the actual truth of the matter. And indeed, I think it is very clear that the intuition that it is wrong to break a promise is not an R-intuition, as it has clear evolutionary relevancy. Communitarian living requires cooperation and coordination among community members, and keeping promises facilitates continued cooperation whereas breaking promises makes it more difficult. If our ancestors tended to feel some sort of special compulsion towards keeping their promises, it seems as though their communities, in the long run, would do better. As such, I do not think that our moral intuitions regarding promise breaking give us a reason for favoring either a deontological or consequentialist approach, since they are not R-intuitions and thus do not play a role in our reflective equilibrium methodology.

Now, some people might be skeptical about how these kinds of moral intuitions are supposed to be developing evolutionarily. I would like to note that I am not necessarily claiming that it is hardwired into our brains that promise breaking is wrong. It also seems possible that our parents inculcate a lot of our moral intuitions during our childhood. However, I think that such inculcation is just as much of an atavistic process, since the types of things we teach our children are likely to have come from things that we were taught by adults that were influential in our lives, and the same goes for those influential adults, and so on. Where did these didactic compulsions about promises come from in the first place? I would point to human ancestry, which found certain types of actions useful to their communitarian lifestyle, and either developed a physiological compulsion towards promise keeping or instilled a psychological one during their
offspring’s development (which led to their offspring instilling the same kinds of values in their progeny, and so on). Whichever mechanism drives the presence of particular moral intuitions does not much matter for my general argument against their relevancy—if they develop out of evolutionary usefulness then they track the gene survival norm, not the moral one.

The sort of argument I just gave against the relevancy of our moral intuition that we should-M keep our promises has deep implications for some of the arguments against consequentialist theories. In his “Ethics and Intuitions,” Singer notes that the chief weapon of opponents to utilitarianism involves examples that are intended to show that utilitarian claims clash with near-consensus moral intuitions. For example, consider an argument advanced by Fred Feldman in Chapter 4 of his Introductory Ethics, which I will refer to as the “Desert Island” case:

A grandpa and grandson crash onto a desert island, and the grandpa is approaching death. Before he passes, the grandpa asks to be buried as his final dying wish, and the grandson promises to do so. But the grandson would get more utility by using his grandpa’s body as bait for fishing, so he does that instead. Most people find that the grandson’s actions are morally wrong, contrary to the judgment of a utilitarian.

While I certainly agree that most people would find that it is wrong for the grandson to use the body as bait for fishing, I do not think that their finding is good evidence that it is actually wrong for the grandson to do so. People’s judgments regarding this case are generally influenced by their intuition that it is wrong to break promises, and since this intuition is not reliable, neither is the ensuing judgment that is based on this intuition. I also think that the strength of the typical intuitive reaction to this case helps my general argument against the reliability of moral intuitions. If our intuition that it is wrong to break promises were evolutionarily engendered, we would expect to find a stronger compulsion against promise breaking in cases that involve people like family members or loved ones than in cases that
involve acquaintances, strangers, or even enemies, since promise keeping is going to be more useful (in evolutionary terms) when it involves individuals that are closer to us than with less distant members of our communities, or even members of other communities.

Imagine another case, which I’ll call the office case.

Andy, Gabe, and Erin all work together and are not related. Andy and Erin used to date, but they broke up about a year ago. Gabe and Erin have been dating for the last six months, until about two days ago, when Erin broke things off with him as well. Gabe is very distraught, and nearly inconsolable. He corners Andy in a conference room, and pleads with him to not get back with Erin. Andy promises Gabe that he and Erin will never date again. A few months pass, and Andy realizes that Erin really is the girl for him, and they get back together, contrary to the wishes of Gabe.

I would bet that most of us do not find the act of promise breaking in the office case to be as bad as that in the desert island case. I propose that this is for a few different reasons. One is that the promise breaking act in the office case does not involve a family member, but rather mere acquaintances, and so the intuitive compulsion we have against the breaking of the promise is not as strong (which matches with my earlier prediction, and lends credence to the evolutionary explanation for our intuitions about promises). The other is that there is no dead body that is being fished with in the office case, an idea that itself creates a feeling of revulsion. But our revulsive sentiments are also evolutionarily engendered; they tend to relate to things that would have direct negative impacts on our health such as consuming human bodies, feces, or spoiled foods. Both of these reasons indicate that our stronger intuitive reaction in the desert island case is due to differences that have a clear relation to our evolutionary history, further indicating that our intuitive reactions to these cases are grounded in our inherited traditions or traits.

The fact that our moral intuitions about promises are unreliable indicators of moral truth leaves a significant challenge for many deontologists: they now need to motivate the existence of
duties against promise breaking without appealing to that intuition, because we should-E not believe that a duty exists without some good reasons for thinking that it exists. Of course, not all deontologists believe that there is such a duty against promise breaking, but for now we will restrict this argument to those who do. I do not think that they can motivate the existence of a duty against promise breaking adequately because they cannot appeal to the consequences of promise breaking (they want to condemn action-types, not action-consequences). I cannot think of any way to motivate claims of the form “actions of type X are wrong” without claiming that they produce bad consequences or without pointing to a particular moral intuition involving actions of that type.

So thus far, we have declined belief on the claim that, as an action-type, promise breaking is wrong. Still, it may very well be that some instances of promise breaking are wrong because of their consequences (as I will argue later). But before we get there, I want to make a general argument against belief in any deontological theory.

There are many action-types that we intuitively find to be wrong. Murder, lying, breaking promises, cheating, physical violence, and incest are all good examples. However, there are also many theoretical situations in which we would tend to think that a person should perform such actions. Imagine any sort of “doomsday” scenario in which the Earth gets blown up unless Dwight murders Jim, or lies to Jim, or cheats on Angela, or commits incest with his mother. In these kinds of scenarios, most people will think that Dwight ought to do what will save the world; namely, murder, lie, cheat, or commit incest. All deontological theories have duties against performing some of these kinds of actions, which is to say that they think we ought not perform them. Either they construe these duties to be absolute (i.e. we should-M never
violate these duties), or they construe them as capable of being overridden by other, more stringent duties. I will first argue against what I term absolute deontological ethics, which construes these duties to be absolute. Later, I will argue against “prima facie deontology,” which involves duties that are capable of being overridden by more stringent duties.

Clearly, absolute deontological ethics takes the unintuitive position when it comes to these doomsday scenarios—that Dwight should-M not murder, lie, cheat, or commit incest in order to save the world. Sure, an absolute deontological ethics might involve modified (but still absolute) duties like “it is wrong to murder someone unless by doing so you save the world,” but then these duties are either being modified in light of various consequences (which a deontologist should-E not be doing), or else we are now operating in the realm of prima facie duties, which I will argue against later on, and so I will not consider these “modified” absolute duties as part of my argument against absolute deontological ethics.

However, even if absolute deontological ethics provides us with unintuitive results in some of these doomsday scenarios, the mere fact that a system of morality conflicts with our moral intuitions—even strong ones—is insufficient for rejecting it. If we are to reject absolute deontological ethics, we need to show that it conflicts with R-intuitions (or rational judgments, which are ultimately grounded in R-intuitions), or else demonstrate that there are no good reasons for believing it and then provide a positive argument in favor of another conception of morality. Here, I will do the former; I think that the R-intuitions that absolute deontological ethics conflicts with are the universal intuition and the intuition that if something is bad, more of it is worse. Consider the following case:

---

1 I will not argue extensively that this intuition is an R-intuition, but I think that it is one. It has no clear role to play in human beings’ evolutionary story, and most people would certainly intuit that it is true.
Angela is in a car. In the seat next to her is Robert, a (State) Senator. They are driving towards a political rally where Robert is going to tell fifty lies to a crowd in order to help his reelection prospects. Angela is the only person that knows the directions to the rally, and if Robert does not make the correct turn ahead then he will be late and someone else will speak instead of him (and this person will tell no lies). When Robert asks Angela which way to turn to get to the rally, Angela thinks about the decision in front of her. Either she can tell one lie and prevent Robert from telling more lies to more people, or she can tell the truth and thus produce a state of affairs in which many more lies are told. Angela has no other options, and cannot prevent the (State) Senator from telling lies by any other means. Angela has no interests in this situation beyond doing what she thinks is best for the world; she wants to do whatever is morally right, as far as everyone’s interests are concerned. She also believes that lying is morally wrong, and that it is bad for everyone in the world when the people in it lie.

Let’s call one possible decision of Angela’s the Deontic Decision, and that’s where Angela decides to tell the truth and Robert subsequently tells fifty lies. The other we will call the Pants On Fire Decision, and that’s where Angela tells a lie and Robert misses his engagement at the rally. If Angela makes the Deontic Decision, her reasoning seems faulty for one reason: if lying is morally wrong, then why would morality dictate that we cannot lie even if that lie is the only way to prevent more lying?\(^1\) I propose that the absolute deontologist can give only two replies to this question, and each answer conflicts with either the universal intuition or the R-intuition that if something is bad, more of it is worse.

There are just two potential morally relevant differences between the Deontic Decision and the Pants On Fire Decision to which a deontologist can point to make their case. The first is that one decision will produce fifty instances of a violative action-type while the other produces only one such instance; this is a difference of clear moral relevancy. But if the absolute deontologist points to this difference as an explanation for why it is worse (in his view) for

---

\(^1\) Samuel Scheffler discussed a similar objection to deontology in his book, *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. He noted that he finds it paradoxical that it may be wrong to do A even if that is the only way to stop several others from doing A.
Angela to tell the one lie, then he clearly lies in conflict with the R-intuition that if something is bad, more of it is worse. So he has to turn to the other possible morally relevant difference between the Decisions, and note that in Pants On Fire Angela tells a lie herself, but in the Deontic Decision she does not tell a lie herself; it is other people that are performing the violative action. But this explanation conflicts with the universal intuition—it only makes sense if there is a reason to think, all things equal, that it is better for other people to do a bad action than it is for you to do that same bad action (even if they are doing more of that bad action). If the absolute deontologist points to this difference to explain why Angela should-M make the Deontic Decision, then they must think that the actions we undertake ourselves are inherently special when compared to the actions of others, just because we are the ones performing them. Since this thinking runs into conflict with rational judgments that are grounded in the universal intuition, it should-E be rejected under our method of reflective equilibrium.

So no explanation that the absolute deontologist can give will keep his ethics from violating at least one R-intuition. This argument, which was restricted to the absolute duty against lying, can be generalized to any candidate bad action that absolute duties forbid us to engage in (let’s call the candidate action ‘A’). If it is morally bad for people to do A, then why is it worse to do A to prevent more instantiations of people doing A than it is to not do A and produce more instantiations of people doing A? No matter the action-type selected, the absolute deontologist will encounter the same problem: in attempting to answer this question, he will invariably come into conflict with either the universal intuition or the R-intuition that if something is bad, more of it is worse. Thus, I find that under our method of reflective equilibrium we should-E reject absolute deontological ethics.
Since absolute deontological ethics is off-limits, there is only one system of ethics left that a committed deontologist can turn to: prima facie deontology. A prima facie system of deontology can accommodate the problem discussed earlier by simply claiming, for example, that we have a prima facie duty to not lie, but this duty is capable of being overridden by another duty against facilitating the production of lies by others. It can also provide more intuitive answers to the “doomsday” scenarios by claiming that Dwight has a prima facie duty not to murder, lie, commit incest, etc., but he also has a more stringent duty against allowing the world to be destroyed. Thus, a prima facie deontologist can coherently claim that Dwight should-M act to save the world in these kinds of scenarios.

One difficulty I have with such a picture of morality is applicable to all deontological theories: it really seems as though the reason that things like lying or murdering are wrong lies in the effects that these actions have on other moral agents. Take, for instance, the intuitive proposition that some lies are worse than others and some murders are worse than others, even when there are no other violative action-types present (i.e. there are some murders that are worse than other murders, just based on the presence of other contextual facts that do not themselves add up to another kind of violative action). This proposition is incomprehensible under any deontological system, since actions are wrong only in relation to how they violate duties, and yet the consequences of actions seem to be entirely relevant in determining whether an act is better or worse than another of the same type. As an example, contrast Dwight lying to Jim about where Jim’s child is just to make him anxious, with Dwight lying to Jim to keep his surprise

---

k This qualification is necessary to sidestep the objection that deontological theories can make coherent sense of the idea that some actions are worse than other actions of the same type as long as the latter involve more duty violations…for instance, they could claim that one act of murder is worse than another act of murder if the former act also involved an instance of promise-breaking and the latter did not.
birthday party a secret. Dwight’s lying in one scenario certainly seems worse than his lying in the other, but he violates the same duty in both scenarios and there are no other duties being violated; hence, if Dwight’s lying really was worse, it must be in virtue of the consequences created by his act.

A prima facie deontologist has two possible lines of response here. One is to modify his “list” of prima facie duties to include duties that condemn lying to people to make them anxious and to exclude duties that condemn lying to people to facilitate birthday surprises (or to include such duties but make them less stringent). Such a modification would provide the necessary framework for claiming that one of Dwight’s lies was worse than the other. However, the modification seems ad hoc at best. Furthermore, consider the fact that we might come up with cases like Dwight’s lies for many other purported duties. If we were to do so, then the number of modifications we might demand of the prima facie deontologist, if he wishes to build an intuitive theory, extends greatly. Additionally, since our deontologist would need to motivate the existence of every one of his various duties that his theory proposes (since we should-E not believe that a duty exists unless we have good reason(s) for believing that it exists), he will have a serious problem motivating his theory as a whole if he makes all of these modifications. Ultimately, I think that his project will fail because the deontologist will be unable to sufficiently motivate the existence of all of his duties, as I will argue later.

The other response is to simply deny that it matters how we tend to distinguish between Dwight’s lies and claim that the lies were both equally wrong, despite our intuitions to the contrary. Of course, the deontologist would also want to make this same claim with regards to other kinds of duties, since we can come up with many other cases in the same vein of Dwight’s
lies. Ultimately, this response is also going to severely undermine the deontologist’s ability to motivate the existence of his duties (again, as we will see later).

There are other theoretical concerns that apply more specifically to prima facie deontology. If an ethical system claims that some duties override others, the system is worthless unless it can also tell us which of those duties are more stringent and which are less so. While there are clear cases (at least intuitively speaking) such as the duty against lying seeming less stringent than the duty against murdering, there are many more unclear ones—is incest as bad as stealing? Is being unfaithful as bad as mugging somebody? Of course, the prima facie deontologist could just make a list of all his duties, ranked by their level of stringency… but any reader of his list would want to know why certain duties are ranked above others; what is it that really makes (for example) lying worse than breaking a promise? The prima facie deontologist cannot point to the consequences of such actions as an explanation, but in making such a list he also cannot point to our intuitions—as no such list is going to be able to make distinctions between all of the candidate wrong action-types in a way such that all the distinctions have widespread intuitive support.

So, a deontologist of whatever stripe needs to both motivate the existence of his various duties and then explain what makes some duty violations worse than other duty violations. Since we have already rejected absolute deontology, we will just focus on how the prima facie deontologist might meet these challenges.

Prima facie deontologists are going to have to draw on our primitive, unreliable moral intuitions to motivate the existence of duties. They cannot point to the consequences of our actions. Perhaps they can derive their ethical systems from R-intuitions, but it is not clear to me
how they might do so (Kant’s derivation of the categorical imperative comes close in terms of his general methodology, but is obviously false), so I will leave that possibility for another to explore; for me, it will suffice if we can derive one ethical system purely from R-Ideas.\(^1\) And in fact, it seems as though many traditional duties do indeed come from our evolutionary impulses.

Consider a partial list of W. D. Ross’s prima facie duties from “The Right and the Good.”\(^8\) Keeping one’s promises and not engaging in deception, making up for wrongful actions, and not harming others are all on his “list” of prima facie duties. Intuitively, all of these are attractive duties: in typical scenarios, we certainly want people to keep their promises and often think it is wrong when people fail to do so (especially when their promises are made to us!). But these are all A-intuitions, which is to say that there are credible underling evolutionary explanations for why we have them; communities that believe and enforce these kinds of prima facie duties tend to collaborate better and persist longer than communities that do not. I already argued, earlier in this section, that our intuitions about promise breaking are A-intuitions; as such, it is impossible to say whether our intuitions regarding promise breaking cut at the very joints of moral truth or if they are mere fantasies selected for by the process of evolution. Hence, this A-intuition (and all others) should-E be construed as unreliable. The other duties listed have similar explanations behind them: communities that inculcate peacemaking by means of reparative actions tend to be more supporting and tight-knit; communities that have a prima facie condemnation for violence (i.e. that tend to disavow violence in cases where there are no overriding factors which would more understandably, perhaps even justifiably, drive one to commit violence, such as a rival faction’s presence threatening family members or the need for territory) are less likely to be torn apart by internal violence.

\(^1\) Additionally, it would be quite problematic for my method of reflective equilibrium if we were able to derive two.
Since the prima facie deontologist cannot point to consequences, and R-intuitions do not lend themselves toward motivating the existence of his duties, he is forced to rely on our moral intuitions more generally and so abandons the method of reflective equilibrium. Not only is this a problem given the general unreliability of our moral intuitions, but it also puts our prima facie deontologist into a corner with regards to his earlier explanations about what makes some actions worse than others of the same type. He wants his theories to be highly intuitive… more intuitive, in fact, than consequentialist theories. So for Dwight’s lies (and any other case we might come up with that leads to a prevalent intuition that one instance of an action of type A is worse than another instance of an action of type A), he is going to have to modify his list of duties, as he cannot coherently claim that our intuitions about these cases are merely false—which was his only other option—since he needs his theory to comport with our general moral intuitions. He will also need to justify the existence of these duties, and lack of such duties in other cases, purely by appeal to our intuitions. This will be difficult, especially when it comes to cases like Trolley and Footbridge. If the prima facie deontologist wishes his theory to match our moral intuitions (which itself already seems like a deeply flawed project), he needs to find a morally relevant difference between those cases so that he can build that difference into his list of duties (as that is necessary for his theory to justifiably claim that the act of “killing-one-to-save-five” in Footbridge is morally worse than the apparently identical [in moral terms] act in Trolley). I will leave that challenge open for a prima facie deontologist to attempt, if he is still inclined to do so in light of the epistemic worries involved in his abandonment of the method of reflective equilibrium.

Furthermore, any system of ethics that is based entirely on general moral intuitions is doomed to fail one important criterion for deciding between moral theories: that they be action
guiding. A good ethical theory is going to give us, at least on a theoretical level, an answer as to what we should-M do across a variety of scenarios. However, general moral intuitions do not converge in enough areas for a theory built entirely around our intuitions to be very action guiding, especially in contentious cases. For instance, is it wrong to get an abortion? If an ethical theory is grounded in our general intuitions, then I do not see how it can possibly give us an answer to this question (at least not at this point in time) given the widespread intuitive disagreement present on the issue; but of course, given the objectivity of morality, there is an answer to this question.\(^m\) The failure of such theories to answer these questions counts further against all theories that are grounded in general moral intuitions. (Of course, some deontological theories do provide answers to these kinds of questions—I would just argue that such theories are even more bereft of good reasons motivating belief in them, for their proponents are not grounding their theory’s justification in widespread, near-consensus intuitive support, but rather just in terms of comportment with *individual or factional* moral intuitions—which are even more unreliable than moral intuitions taken generally).\(^n\)

So to conclude this section, I would note that we have many good reasons to want to steer away from a deontological approach to ethics. However, these reasons are not decisive unless we also have good reasons that positively favor taking a different approach. Since, as I will argue presently, we have reasons for believing that the consequentialist approach is preferable under the method of reflective equilibrium, I conclude that we should-E reject deontology.

\(^m\) Although it might not be a yes or no answer. The existence of a true “yes and no” answer that specifies certain contexts or conditions under which abortion is permissible and impermissible would satisfy the demands of moral objectivity here.

\(^n\) Imagine: if our general moral intuitions track the gene survival norm, our individualized intuitions are going to track the “random-genetic-mutation norm.”
IV. OR CONSEQUENTIALISM?

As noted in the previous section, it is an R-intuition that if something is bad, then more of it is worse. I would also submit that the following is an R-idea: some actions are worse than others of the same type (and it follows that ‘some actions are worse than others’ is also an R-idea). These R-ideas posed serious problems for a deontological approach to ethics, both for an absolute deontologist and for a prima facie deontologist, due to the prevalence of cases like Dwight’s lies (and their subsequent need to create duties that differentiate between actions in pairs of cases that seem to involve actions of the same type, with one being clearly worse than the other). Furthermore, there are larger worries about any ethics project that abandons the method of reflective equilibrium for an approach that matches the truth of moral theories with their overall comportment with our general (and unreliable) moral intuitions.

Since comportment with R-ideas is an important criterion under the method of reflective equilibrium, an ethical theory that can, on the theoretical level, both describe and coherently explain why certain actions are worse than others of the same type is going to balance out more favorably than deontology. The same goes for the R-intuition that if something is bad, then more of it is worse. I will argue that consequentialist theories can just as smoothly meet the challenges posed by these R-ideas as the ideal prima facie deontological theory can, and that they have a firmer theoretical foundation on which to stand. (The ideal prima facie deontological theory is the one that best comports with these R-ideas, and it will be the one that contains an extensive list of prima facie duties that distinguish between actions in cases like Dwight’s lies. However, such a theory comports with these R-ideas at the cost of a reasonable theoretical foundation, since many of the duties on “the list” are going to be completely unintuitive, and their rank-ordering in terms of stringency will also be both unintuitive and ad hoc). Additionally,
consequentialist theories can also explain why many of the “traditional” deontological duties do exist (in a sense), which actually is somewhat important, as I shall argue presently. If all of this is true, then it follows that we should favor consequentialism over deontology under the method of reflective equilibrium.

Does it matter whether a theory comports with our general moral intuitions? Even though I construe our general moral intuitions to be unreliable, I still think that the answer is yes. I will argue for two claims here. Claim 1: all things equal, theories that better comport with our general moral intuitions should be favored under the method of reflective equilibrium. Claim 2: if that “favored” theory were to conflict with an R-intuition and another theory does not do so, then that “favored” theory should no longer be favored.

Claim 1 is true because R-intuitions are reliable but hard to discover. The line between an A-intuition and a B-intuition is a vague one, as is the line between a near-consensus intuition and a more divergent intuition. Since these lines are vague, I have elected to just isolate clear instances of R-intuitions from which to build my arguments. However, this “cautious” approach means that the general pool of moral intuitions is likely to include some reliable intuitions that I will fail to unearth. Furthermore, I made the distinction between R-intuitions and the rest of our moral intuitions just because R-intuitions are the only sorts of intuitions that I have good reason to think are reliable, but there might be other kinds of moral intuitions that are reliable. For instance, there may be some near-consensus intuitions that are not evolutionarily engendered and yet there seems like there is a good evolutionary explanation for why they exist; these intuitions could be reliable and yet are not R-intuitions. I will call these two kinds of reliable intuitions (those neglected due to my cautious approach, and those ignored because they are A-intuitions but aren’t really evolutionarily engendered) “hidden intuitions.” Given the probable existence of
hidden intuitions, it appears that general comportment with our moral intuitions should count in favor of the truth of a moral theory, since it makes it more likely that such a theory comports with some hidden intuitions.

Claim 2 is true because we still need to be cautious about the unreliability of most moral intuitions. Since we have good reasons for thinking that R-intuitions are reliable, comportment with them should be privileged over comportment with the general pool of moral intuitions. The amount of evidence that a high level of comportment with the general pool of moral intuitions provides us for thinking that a candidate theory is true is unclear, because we do not have any indication about the number of hidden intuitions vis-à-vis the number of unreliable general moral intuitions. Given these sorts of epistemic concerns, I do not think that it would be reasonable to privilege a theory with a high level of comportment with our general moral intuitions that conflicts with at least one R-intuition over a theory that does not conflict with any R-intuitions and yet does not have a high level of comportment with our general moral intuitions.

It is because of claim 1 that I think that it is of some value that a consequentialist theory be able to explain the “existence” (in a sense) of many traditional prima facie duties, given that most people intuit that there is at least some truth involved in such duties. And in fact, I think that consequentialist theories can do so. I will restrict myself to explaining how, under utilitarianism, a sort of prima facie duty against breaking promises, lying, murdering, and stealing exists, but I think that my explanation is similar to one that can be given by many other kinds of consequentialist theories.

Utilitarianism holds that actions are morally right iff they are available and produce at least as much utility as any other available action. I think that, more often than not, actions that are instances of breaking promises, lying, murdering, and stealing will not produce such a
“maximal” utility level. Imagine that we are firm believers in utilitarianism, and we hear that Dwight has broken a promise, but we do not have any other information about what Dwight did. Now imagine that we are forced to make a moral assessment of Dwight’s action: we have to either say that Dwight acted rightly or that he acted wrongly, all without learning anything more about what Dwight actually did. Since it is true that more often than not, actions that are instances of breaking promises produce a utility level that is less than maximal, our best guess would be to claim that what Dwight did was wrong. (This line of reasoning can obviously be extended to a case in which we hear that Dwight told a lie, murdered someone, or stole something.) Thus, a proponent of utilitarianism can reasonably believe that it is generally wrong to perform actions that are instances of promise breaking, which is a belief that is similar to the belief that we have a prima facie duty against breaking promises. In that sense, a consequentialist theory can be said to comport to with our intuitions about the existence of certain prima face duties, which counts at least slightly in favor of such consequentialist theories.

Next, I will argue that consequentialist theories can smoothly accommodate the R-idea that if something is bad, then more of it is worse, and the R-idea that some actions are worse than others of the same type. Consequentialist theories take particular consequences of actions to have moral relevance, and suggest that the rightness or wrongness of actions is determined entirely by the sort of consequences that they produce. For instance, a consequentialist theory might claim that the morally relevant consequences of actions are those that increase or decrease the total amount of knowledge in the world. Such a theory could easily accommodate the R-idea that if something is bad, then more of it is worse by claiming that producing a net loss of knowledge in the world is bad, and the more overall knowledge lost, the worse. Furthermore, such a theory can easily explain why it is worse for more knowledge to be lost: because losing
knowledge is bad, so the more knowledge that is lost, then the more badness that exists in the world. (Of course, such a theory needs to motivate why losing knowledge is bad, but we can easily see how this R-idea would fit into such a theory, which is what I’m going for.)

Utilitarianism is another consequentialist theory that can easily accommodate the R-idea that if something is bad, then more of it is worse. Utilitarianism takes the “bad” consequence of actions to be pain. It is obviously coherent for a proponent of utilitarianism to claim that if pain is bad, then more pain is worse; hence, the theory easily accommodates this R-idea.

Utilitarianism also easily accommodates with the R-idea that some actions are worse than others of the same type, and it can more smoothly (than deontological theories) provide an explanation for why particular actions are worse than others. Let’s briefly return to the case of Dwight’s Lies. Dwight tells two lies to Jim, but it seems as though one lie was clearly worse than the other, even though both lies violate just one duty: the duty against lying. Utilitarianism would analyze the two lies differently: one lie produced anxiety in Jim (since Dwight lied about the location of Jim’s child), and since anxiety is a negative emotional state which fits under the umbrella term ‘pain,’ it is fair to say that Dwight’s lie produced net pain in the world. The other lie facilitated Jim’s birthday surprise, which we might stipulate that he thoroughly enjoyed, and so we can claim that this lie produced net happiness in the world; hence, since one lie produced more overall badness than the other lie, it was a worse lie. So, not only can this consequentialist theory give us the sort of results we would expect with regards to the R-idea that some actions are worse than others, it can also give us a mechanism for determining whether any candidate action was better or worse than another: simply look at the differences in overall utility produced by the two actions, and if one action produced less overall utility than the other, then it was a worse action.
There are other consequentialist theories that both comport with and can explain the two R-ideas mentioned here, but I will not talk about any of those—it suffices for this argument that I have demonstrated that some consequentialist theories fully comport with these two R-intuitions.

Next, I will argue that these consequentialist theories have a firmer theoretical foundation on which to stand than the only kind of deontological theory that comports with both of these R-ideas (the “extensive list” prima facie duty theory). The argument here is simple: a consequentialist theory merely needs to motivate the belief that whatever effects of actions it construes as having moral relevance (be it the production of utility, knowledge, or something else) are indeed good or bad effects. The extensive list prima facie duty theory, on the other hand, needs to motivate belief in the existence of every single one of its purported prima facie duties, which is not only a much bigger task but it also a much more difficult one (especially if some of these duties need to draw distinctions between the actions involved in Trolley and Footbridge, where any potential distinction seems doomed to moral irrelevancy). Because some of these effects are clearly morally relevant (the most obvious being the production of utility), it follows that the theoretical foundation for some consequentialist theories is much firmer than that for the ideal prima facie duty theory.

In this section, I have demonstrated that consequentialist theories can comport with the following two R-ideas: if something is bad then more of it is worse, and some actions are worse than others of the same type. I have argued that consequentialist theories stand on a firmer theoretical foundation than the only deontological theory that comports with both of these R-ideas. I would submit that, under the method of reflective equilibrium, this demonstrates conclusively that consequentialism should-E be preferred to deontology. The question, at this
point, becomes which sort of consequentialism underlies the correct moral theory—but that is a question for another project.

Works Cited