

# A Fault Line in Ethical Theory

## Consequentialism, Deontic Constraints, and the Prisoner's Dilemma

Shyam Nair

### 0. Introduction

A venerable idea in the history of moral philosophy is that *central* among the normative notions is the notion of goodness or value. This idea, which can be found at least as early as 1903 in G.E. Moore's *Principia Ethica*, claims that goodness is central in that all other normative notions can be explained in terms of it. Moore's approach and the dominant approach to the project of explaining the normative notions of rightness and what we have reason to do is the *consequentialist program*. While the most famous element of this program is the consequentialist moral theory that says an act is right (i.e., permissible) just in case the outcome of that act is at least as good as the outcome of any alternative act, the consequentialist program in general hopes to similarly explain the notion of a reason in terms of goodness. So the consequentialist theory of practical reason says that an agent has more reason to do some act *a* than to do some act *b* just in case the outcome of doing *a* is better than the outcome of doing *b*.

A consequentialist theory of either of these sorts, we are told, is incompatible with cases of *deontic constraints*. A case of *deontic constraints* is supposed to be a case where an act is wrong (or one that there is most reason to not do) even though performing that act will prevent more acts of the same morally (or practically) relevant type from being performed. What we are told is that a fundamental choice point in ethical theory is whether there are cases of deontic constraints: there are consequentialist theories that cannot allow for the existence of these cases and there are so-called "agent-relative" theories that can allow for their existence. In this way, cases of deontic constraints are thought to form a kind of fault line in ethical theory separating consequentialist theories from other theories. And indeed, the traditional wisdom has been that this fault line is of great theoretical importance. More carefully, though philosophers have not often explicitly announced that this fault line is of great importance or explicitly explained how it is of great importance, the standard practice of ethical theorists (which we will look at in detail shortly) has been to treat this fault line as playing a number of important theoretical roles.<sup>1</sup>

---

<sup>1</sup> To my mind, Scheffler 2003 (esp ch. 4) is the classic example of this wisdom in action. But this idea is expressed or assumed in slightly different terminology in other classic sources as well, see Rawls 1971: 26, Smart and Williams 1973: 89, Nozick 1974:

Unfortunately, careful study of the literature in moral philosophy reveals certain results that are hard to reconcile with this tidy picture of the importance of cases of deontic constraints. A particularly striking result comes from “Act and Value”, a clever paper by Graham Oddie and Peter Milne. In this paper, Oddie and Milne begin by making several relatively uncontroversial assumptions about the nature of goodness (e.g., that it determines a transitive ordering on outcomes) and the commitments of consequentialism (e.g., that it is a maximizing theory). They then prove what they call a “representation result” from these assumptions. This result is significant for our purposes because it follows trivially from it that consequentialism *can* accommodate cases of deontic constraints. So if Oddie and Milne are right, the idea that cases of deontic constraints form a fault line is not an important result that was established by moral philosophers in the 20<sup>th</sup> century; instead, it is a demonstrably false dogma that these philosophers naively accept.

My aim in this paper is to argue that both the traditional wisdom and Oddie and Milne are mistaken. Oddie and Milne are mistaken because cases of deontic constraint do in fact form a fault line in ethical theory. The traditional wisdom is mistaken because this fault line does not have the importance claimed for it. Instead, I will show that a new argument and set of assumptions is needed in order to establish the fault line and appropriately circumscribe its importance in ethical theory.

I will start working my way toward this aim by looking more closely at the traditional wisdom and why it fails (§1). Then I will turn to developing my account of the fault line in ethical theory and its importance (§2). Finally I will show that a diagnosis of where Oddie and Milne go wrong follows as a corollary of this new account (§3). An important upshot will be that discussions of cases of deontic constraints would do best to focus on the account of the nature and importance of these cases that identify in this paper rather than continuing to work with the mistaken traditional picture when assessing whether the correct ethical theory should allow for such cases.

## 1. The Importance of the Traditional Wisdom and Its Problems

Let’s begin by considering the issue of deontic constraints more closely. In the first instance, philosophers are concerned about deontic constraints because common sense moral and practical thinking recognizes their existence. To illustrate, consider the case of Mary:

Mary faces a choice of whether or not to murder some innocent person. If Mary does not murder this innocent person, Sally and Tim will each murder a different innocent person. But if Mary does murder this person, she will thereby prevent each of Sally’s and Tim’s murders.

---

28-33, and Darwall 1986: §1. The idea and the cases that motivate it are also often used to illustrate distinction between agent-neutral and agent-relative, see Nagel 1979: ch. 10 and Parfit 1984: 27.

Our common sense moral thought suggests that it would be wrong (i.e., not right) for Mary to murder even though her murdering would prevent more acts of the same morally relevant type (two murders) and that it would be right for Mary to not murder even though this will allow more acts of the same morally relevant type to occur. Similarly, common sense practical thought suggests that Mary has better reason to refrain from murdering than to murder even though murdering would prevent more acts of the same practically relevant type from occurring. And while this particular case involves murders, common sense moral and practical thinking recognizes many other cases as well. For example, common sense similarly recognizes that it would be wrong to break a promise and wrong to fail to help your loved ones even if by doing so others would keep their promise and help their loved ones.

By considering cases like this, philosophers arrived at the technical definition of a case of deontic constraints that I mentioned at the beginning of this paper: a deontic constraint *in this technical sense* is a case where an act is wrong even though performing that act will prevent more acts of the same morally relevant type from being performed.<sup>2</sup> While I think this technical definition is reasonable, we should be careful about how we use it. In particular, some theorists may want to use their theory to inform what counts as ‘relevant’. If they do, they might accept the existence of these common sense cases but deny that any of them meet the technical definition of a case of deontic constraints because their theory claims that there is a relevant difference between the acts.

But of course what is important about cases of deontic constraints is not that they conform to this technical definition but rather that they are an important part of common sense moral and practical thought. For this reason, we should use ‘deontic constraints’ as a tag for the class of cases like Mar’s case that we pretheoretically recognize. So officially the question that I am pursuing is whether the consequentialist program is compatible with the existence of deontic constraints in this sense.

For this reason and for the purposes of simplicity and concreteness, we will use the case of Mary as our representative of cases of deontic constraints in morality and practical reason. And throughout this section, we will further simplify things by exclusively discussing morality. It should however be easy to see how our discussion generalizes.

What I want to argue in this section is that the traditional wisdom about the importance of the fault line in ethical theory formed by cases of deontic constraints form is mistaken. Let’s begin then by looking at two of the important roles that the traditional wisdom claims (or, again more carefully, the standard philosophical practice of ethical theorists assumes) this fault line plays.

---

<sup>2</sup> Cf. Scheffler 2003: 84

## 1.1 Depth

To see the first role, we must distinguish cases of deontic constraints from cases of *deontic options*. Cases of *deontic options* are cases where it is right or permitted for an agent to do each of two acts even though the outcome of one of the acts is better than the outcome of the other. These cases are thought to arise when the act that has the worse outcome is especially important to the projects of the agent.

Cases of deontic options are importantly different than cases of deontic constraints. To see what is different about them, we must note two logical features of consequentialism. First, consequentialism entails the following claim:

*Compelling Idea*: If an act has the best outcome (i.e., the outcome that outranks the outcomes of all the others acts), then it is right.<sup>3</sup>

Second, consequentialism is equivalent to the conjunction of the following two theses:

*Left-to-Right*: if an act is right, then it has an outcome that is not worse than the outcome of any alternative act.

*Right-to-Left*: if an act has an outcome that is not worse than the outcome of any alternative act, then it is right.

The difference between cases of deontic constraints and cases of deontic options concerns which of these theses they are incompatible with.

In particular, cases of deontic options are only incompatible with *Left-to-Right*. In these cases, it is right for the agent to do some act that furthers her personal projects even though it has an outcome that is worse than the outcome of another available act. However, these cases are not incompatible with the *Right-to-Left* because the agent is still permitted to do the act that has the better outcome. And for this reason, these cases are compatible with *Compelling Idea* as well. To label the fact that these cases are not incompatible with all three of these theses, I will say that these cases form a *shallow* fault line in ethical theory.

But according to the traditional wisdom, cases of deontic constraints are supposed to form a *deep* fault line in ethical theory in the sense that they are supposed to be incompatible with all three of these theses. It is supposed to be wrong to murder and right to not murder even though the outcome murdering is better than the outcome of any other available act. If this supposition is correct, the act with the best outcome is not right and this shows that these cases are incompatible with *Compelling Idea* and *Right-to-Left*. And if this supposition is correct, the right act has a worse outcome than another available act and this shows that these cases are also incompatible with *Left-to-Right*.

---

<sup>3</sup> See Foot 1985, Dreier 1993, Portmore 2005.

Noting this supposed difference between deontic options and deontic constraints is theoretically important because philosophers have thought that cases of deontic constraints are paradoxical in a way that cases of deontic options are not. In particular, on at least one reading of the paradox of deontology, recognizing cases of deontic constraints leads to the paradox partly in virtue of these cases being incompatible with *Compelling Idea*.<sup>4</sup> Thus, the first important role the traditional wisdom claims for the fault line is that it is a *deep* rather than shallow fault line in ethical theory.

## 1.2 Neutrality

To see the second role, it helps to consider a different kind of case that might seem to be incompatible with consequentialism. For example, suppose that I could give \$50 to John or to Bill and that Bill would benefit slightly more from the \$50 than John. But suppose further that I have promised John the \$50. Most of us believe that it would be wrong to give the money to Bill and right to give the money to John. And this case might seem like trouble for the consequentialists because it may seem that the outcome of giving Bill the money is best because he would benefit more.

But this troublesome case for the consequentialist (insofar as it is one) is supposed to be importantly different from the trouble posed by deontic constraints. This is because consequentialists have a ready reply to this problem: While this case may be a problem if we have a simplistic theory of goodness or *axiology* according to which, say, pleasure is the only thing that is good, it is not a problem if we abandon this hedonist axiology. If, instead, we have a pluralist axiology according to which things other than pleasure such as knowledge, virtue, friendship etc. are good, then we can easily handle the case. The simplest way for us to do so would be to add promise keeping to the list of goods.

More generally, the consequentialist determines what acts are right by a ranking of outcomes according to their goodness. But the consequentialist is not committed to any particular axiology in virtue of accepting consequentialism. And once the consequentialist has moved beyond a monistic axiology according to which, e.g., pleasure is the only good, they are free to think that the ordering of outcomes according to their goodness is determined by pleasure as well as other goods. This gives the consequentialist considerable flexibility to accommodate common sense moral thinking.

Campbell Brown nicely illustrates just how powerful and distressing this strategy can be:

Arguing with a consequentialist can be frustrating. Witness a typical sort of exchange: You—a nonconsequentialist, let's assume—begin with your favorite counterexample. You describe some action, a judge's convicting an innocent man to avert a riot, say, or a doctor's murdering a healthy patient for her organs, which, so you claim, would clearly have the best consequences, yet equally clearly would be greatly immoral. So consequentialism is false, you conclude; sometimes a person ought not to do what would have best consequences.

---

<sup>4</sup> This is one of the lessons of Scheffler 2003. For discussion of the paradox of deontology, see also Foot 1985, Darwall 1986, Bennett 1989, Brooks 1991, and Hurley 1997.

“Not so fast,” comes the consequentialist’s reply. “Your story presupposes a certain account of what makes consequences better or worse, a certain ‘theory of the good,’ as we consequentialists like to say. Consequentialism, however, is not wedded to any such theory. We already knew that combining consequentialism with some theories of the good would have implausible results; that’s what utilitarianism has taught us. In order to reconcile consequentialism with the view that this action you’ve described is wrong, we need only to find an appropriate theory of the good, one according to which the consequences of this action would not be best. You say you’re concerned about the guy’s rights? No worries; we’ll just build that into your theory of the good. Then you can be a consequentialist too.” (2011: 749-50)

In short, it is hard to find cases that are incompatible with consequentialism once we have recognized that consequentialists can adopt a pluralist axiology. And what is important about the idea that cases of deontic constraints form a fault line is that it tells us cases of deontic constraints are one of these hard to find cases that is incompatible with consequentialism.

In particular, the idea is that even if the only axiological assumptions that we make are structural or logical assumptions about goodness (e.g, that it determines a transitive ordering on outcomes), we can show that consequentialism is incompatible with cases of deontic constraints. In short, it is supposed to be that even if we are *neutral* about which axiology is correct, we can show that cases of deontic constraints are incompatible with consequentialism. And this is the second important role that the traditional wisdom claims for the fault line. It is supposed to be that the fault line can be established on (axiologically) *neutral* grounds. And in fact, one of the main aims of Brown’s article (which we will consider in more detail shortly) is to show how this can be done.<sup>5</sup>

We can further bolster this point by considering a standard diagnosis of why the fault line can be established on neutral grounds. As I am understanding consequentialism, it is consistent for a consequentialist to adopt a pluralist axiology but all consequentialists (in my stipulative usage) are committed to thinking that these different goods together determine a *single* ranking of outcomes according to how good they are. And the standard diagnosis of why consequentialism cannot accommodate cases of deontic constraints is that it appeals to this single ranking.

Douglas Portmore sums this thought up at the outset of his *Commonsense Consequentialism*:

Consequentialism is typically taken to be an agent-neutral theory: a theory that gives each agent the exact same set of aims. Unfortunately, no agent-neutral theory can accommodate all of our common-sense moral intuitions. For instance, no agent-neutral theory can accommodate the intuition that it would be wrong to break a promise even so as to prevent two others from breaking a comparable promise, for the only way to accommodate such an intuition is to give each agent the distinct aim of, say, minimizing his or her own promise-breakings—thus, giving me the aim of minimizing *my* promise-breakings and you the aim of minimizing *your* promise-breaking. (2011: ix)

Portmore here uses the philosophical terminology of “agent-neutral” and cashes this out in terms of the aims of agents. In our terms, we can put the idea like this: In recognizing cases of deontic constraints,

---

<sup>5</sup> See Darwall 1986: §2 for a historical perspective. He says that deontic constraints gained prominence in the 20<sup>th</sup> century because Moore was the first major consequentialist to decisively break with the tradition of adopting a hedonist axiology.

common sense moral thinking suggests an agent ought to have special concern for her actions. For example, Mary ought to be especially concerned about *her* murders. We could, then, for each agent, rank outcomes according to the concerns that agent ought to have. In this way, we would get one ranking of outcomes according to what Mary ought to be concerned about and a perhaps different one according to what Sally ought to be concerned about. Portmore's claim is that consequentialism cannot accommodate cases of deontic constraints because, unlike common sense morality, consequentialism uses a *single* ranking of outcomes according to how good they are to determine what is right and wrong. Thus the standard diagnosis says that consequentialism cannot accommodate cases of deontic constraints simply in virtue of the axiologically neutral fact that it uses only a single ranking.<sup>6</sup>

What's more, the idea that we can establish on neutral grounds that cases of deontic constraints form a fault line has played an important role in moral theorizing. It has led to a new family of theories called agent-relative consequentialism or teleology.<sup>7</sup> These theories are not consequentialist as I am using this term. Instead, they are close cousins to consequentialism. According to these theories, there is not a *single* ordering of outcomes according to how good they are. Instead, there are multiple orderings of outcomes. Each ordering is an ordering of outcomes according to how good they are relative to some agent.

As Portmore's diagnosis suggests and the founding article of this family of theories, Amartya Sen's "Agency and Rights", attests to, one of the primary motivations for the move to agent-relative teleology is to accommodate cases of deontic constraints while sticking close to consequentialism. Evidently, this motivation for these theories is crucially dependent on the idea that we can establish that cases of deontic constraints are incompatible with consequentialism on neutral grounds.<sup>8</sup>

Thus, our discussion in §1.1-1.2 reveals that the traditional wisdom claims (or, again more carefully, the standard practice of moral philosopher implicitly assumes) that cases of deontic constraints are incompatible with all three theses identified earlier and that they are incompatible with these theses no matter which axiology we adopt. In slogan form then, the traditional wisdom is that we can show on *neutral* grounds that cases of deontic constraints form a *deep* fault line.

---

<sup>6</sup> Brook 1991 and Kamm 1996 argue that that is features of patients and not agents that explain the existence of cases of deontic constraints. I do not deny Brook's and Kamm's point. I insist only that if they are right, this is reflected in the ranking of outcomes according to what each agent ought to prefer.

<sup>7</sup> See Sen 1982; Dreier 1993 and 2011; Smith 2003 and 2009; Louise 2004; Portmore 2009 and 2011. For criticism see Regan 1983 and Schroeder 2006 and 2007.

<sup>8</sup> Having an neutral argument is also helpful for two more reasons. First often part of the paradox of deontology invokes the worry that recognizing cases of deontic constraints saddles us with 'intuitions in search of a foundation' (Scheffler 2003: 80). It would be at least dialectically awkward if the only reason we have for thinking that consequentialism cannot provide a foundation for these intuitions is that we have certain intuitions about goodness that cannot themselves be given a foundation. Second it may put the traditional wisdom on surer footing when it comes to various consequentialist perspectives on the role of case based intuitions. For a sampling of the different view see Singer 1974 and 2003, Hare 1980, Kagan 2001, and Portmore, 2011: 111-117.

### 1.3 A Non-Neutral Argument

It was worth spending the time introducing these roles that the traditional wisdom claims this fault line plays because these roles constrains what kinds of arguments we can offer for thinking there really is such a fault line. In this section, I will illustrate this by considering some arguments that fail because they are not neutral in the sense that I just identified.

In Mary's case, it is supposed to be wrong for Mary to murder and right for her to refrain. One way the consequentialist could get this result is by saying that the outcome in which Mary refrains from murdering is better than the outcome in which she murders. Evidently, then, if consequentialism is incompatible with cases of deontic constraints, there must be some reason why she cannot rank the outcomes like this.

One simple idea about why we cannot rank the outcome of Mary's refraining murdering higher than the outcome of her murdering is that it is extremely plausible that the outcome of Mary's murder is better than the outcome of her refraining from murder. The trouble for this simple idea is that it is not neutral. We saw that according to the traditional wisdom, cases of deontic constraints were thought to be incompatible with consequentialism no matter what axiology, however implausible, we chose. But this argument relies on ruling out certain axiologies based on their implausibility.

Similar problems infect more sophisticated arguments suggested by the work of certain philosophers.<sup>9</sup> Though these arguments differ in detail, we will work with the following one:

Grant, for *reductio*, that Mary's not murdering has a better outcome than her murdering. This would mean that Mary's murder is worse than Sally's and Tim's murders. But now suppose that Sally and Mary "switched places". That is, suppose Sally could now murder in order to prevent each of Mary and Tim from murdering. Given that the consequentialist is committed to thinking Mary's murder is worse than Sally's and Tim's murders, the consequentialist is also committed to thinking that Sally's murder is *better* than Mary's and Tim's murders. So the consequentialist is committed to thinking Sally's murdering is not wrong. But there is nothing special about Sally that makes her any different from Mary. It is wrong for each of them to murder even if doing so would prevent two murders.

This argument or some nearby variant of it is, I think, among the best we can find in the literature for the conclusion that the outcome of Mary's murder is better than the outcome of her refraining from murder. Unfortunately, it too fails to be neutral.

To see, this let's work through what it says. A natural way to understand it is as asking us to consider a pair of cases. Letting 'M' stand for 'Mary murders', 'S' stand for 'Sally murders', and 'T' stand for 'Tim murders', we can say that in the first case Mary can realize one of the following two outcomes:

---

<sup>9</sup> It is not exactly clear who is giving this argument and who is giving the simple argument because most philosophers go through this discussion very quickly (though see Scheffler 2003: ch. 4, Darwall 1986: §1, Brook 1991: 1-2). The version presented here most closely resembles the one in Schroeder 2007: §I.B.



$o_1: M, \neg S, \neg T$   
 $o_2: \neg M, S, T$

That is, Mary can murder or not and her murdering will prevent each of Sally and Tim from murdering. In the second of the pair of cases, Sally can realize either of two outcomes

$o_3: S, \neg M, \neg T$   
 $o_4: \neg S, M, T$

That is, Sally can murder or not and her murdering will prevent each of Mary and Tim from murdering.

At first glance, it can seem as though there is no barrier at all to the consequentialist getting the result that it is wrong for Mary to murder and wrong for Sally to murder. Just say  $o_2$  is better than  $o_1$  and say  $o_4$  is better than  $o_3$ .

But advocates of the argument that I have described above have claimed to find a reason for thinking that if you say  $o_2$  is better than  $o_1$ , then you must, contrary to what is desired, say that  $o_3$  is better than  $o_4$ . The argument begins by claiming that the only reason you could have for ranking  $o_2$  higher than  $o_1$  is that you think *Mary's* murders are worse than Sally's and Tim's murders together. But if this is true, then surely, Sally's murder is better than Mary's and Tim's murders taken together. So we have to say the outcome in which Sally murders— $o_3$ —is better than the outcome in which Mary and Tim murder— $o_4$ . And this argument seems axiologically neutral because it does not just rely on our intuitive judgment that it cannot be both that  $o_2$  is better than  $o_1$  and  $o_4$  is better than  $o_3$ . Instead, it shows that if we were to grant that  $o_2$  is better than  $o_1$ , it would follow that  $o_3$  is in fact better than  $o_4$ .

As seductive as this argument is, it rests on a simple mistake. It is just not true that the only reason you could have for ranking  $o_2$  higher than  $o_1$  is that you think *Mary's* murders are especially bad. Instead, the grounds for ranking  $o_2$  higher than  $o_1$  may be something about Mary's circumstances. In particular, if what makes  $o_2$  better than  $o_1$  is that  $o_1$  involves preventing two murders, it would not follow that  $o_3$  is better than  $o_4$ . To the contrary, it would follow that  $o_4$  is better than  $o_3$ .

Now, of course, it may seem simply implausible to think that  $o_2$  is better than  $o_1$ . But, as we saw, it was supposed to be that consequentialism is incompatible with cases of deontic constraints on neutral grounds. Moreover, we can sketch a picture on which this ranking is at least not totally crazy. Distinguish two ways in which someone might fail to murder. She might fail to murder with *dirty hands* or with *clean hands*. We can say her *hands are dirty* in failing to murder when she would have murdered if others had not intervened and would have murdered if other important goods were at stake. This is to be contrasted with failing to murder with *clean hands*. To do this is to fail to murder and for it to be the case that you

would fail to murder even if others did not intervene and even if other goods were at stake.<sup>10</sup> Now suppose our pluralist consequentialist says *dirty handed* refraining from murder is worse than *clean handed* refraining from murder. If such a consequentialist thought clean handed refraining from murder is sufficiently better than dirty handed refraining from murder, she could claim that in the first case, the outcome in which Mary refrains from murdering is better than the outcome in which she murders—the outcome in which she refrains from murdering has one clean handed refraining from murdering and two murders whereas the outcome in which she murders has two dirty handed refrainings from murdering and one murder. And, of course, on analogous grounds she could claim that in the second case, the outcome in which Sally refrains from murdering is better than the outcome in which she murders.

Thus, we have seen that both a simple argument and a more sophisticated argument fail to establish on neutral grounds that cases of deontic constraints are incompatible with consequentialism contrary to the traditional wisdom's claim that the fault line can be established on neutral grounds.

#### 1.4 A Shallow Argument

Having seen that certain arguments fail because they are non-neutral, I now turn to a recent argument due to Campbell Brown that is neutral (2011: 761).<sup>11</sup> The (structure of the) case Brown discusses is slightly different than the one that we have been considering so far. In his case, we are to suppose that Mary can murder and thereby prevent Sally from murdering or refrain from murdering and thereby allow Sally to murder. In symbols:

- (1)  $M \rightarrow \neg S$
- (2)  $\neg M \rightarrow S$

Similarly suppose that in this very same case, Sally can murder and thereby prevent Mary from murdering or Sally can refrain from murdering and thereby allow Mary to murder. In other words, in this case we also have that:

- (3)  $S \rightarrow \neg M$
- (4)  $\neg S \rightarrow M$

So overall, there are only two outcomes that could result no matter what Mary and Sally do:

---

<sup>10</sup> In order to achieve these goods, an agent must act in certain ways not just in the actual world but also in counterfactual possibilities. Though it often does not receive explicit comment and is, therefore, not obvious, I think many pluralist consequentialists are best interpreted as recognizing such goods (e.g. Pettit and Brennan 1986, Railton 1986). But for a particularly explicit example, consider Philip Pettit's account of the good of non-domination (List 2005 and Pettit 2011).

<sup>11</sup> Cf. Portmore 2011: 86-87. This argument also only establishes a shallow fault line.

$o_1: M, \neg S$   
 $o_2: \neg M, S$

In this case, it is supposed to be wrong for Mary to murder and right for her to refrain from murdering. And similarly, it is supposed to be wrong for Sally to murder and right for her to refrain from murdering.

Brown's argument is neutral because he considers all possible ways of ranking  $o_1$  and  $o_2$  and shows that none of them allow the consequentialist to get the desired result in the case. Suppose we said  $o_1$  is better than  $o_2$ . According to consequentialism, it would then be right for Mary to murder and so we would not get the desired result that it is wrong for her to murder. Suppose instead, we said that  $o_2$  is better than  $o_1$ . According to consequentialism, it would then be right for Sally to murder and so we would not get the desired result that it would be wrong for her to murder. Finally suppose we said  $o_1$  and  $o_2$  are equally good. According to consequentialism, it would be right for Mary to murder and so we would not get the desired result that it would be wrong for her to murder. Thus, consequentialism is incompatible with cases of deontic constraints.

This argument is sound and neutral. But it does not suffice to establish that there is a *deep* fault line in ethical theory. According to the traditional wisdom, the fault line was supposed to be *deep* in the sense that these cases are supposed to be incompatible with *Compelling Idea*, *Left-to-Right*, and *Right-to-Left*. This was what is supposed to make cases of deontic constraints importantly different from cases of deontic options and what is supposed to make these cases especially paradoxical.

To see that Brown's argument does not establish a deep fault line, suppose, plausibly, that  $o_1$  and  $o_2$  are equally good. If we adopted this ranking, we can show that cases of deontic constraints are incompatible with *Right-to-Left*—if an act has an outcome that is not worse than the outcome of any alternative act, then it is right—because, e.g., Mary's murdering has an outcome that is not worse than the outcome of any alternative act and yet Mary's murdering is wrong. But we cannot show that they are incompatible with *Left-to-Right*—if an act is right, then it has an outcome that is not worse than the outcome of any alternative act—because the consequent of this conditional is true of each of the acts when  $o_1$  and  $o_2$  are equally good. And we cannot show that they are incompatible with *Compelling Idea*—if an act has the best outcome, then it is right—because the antecedent of this conditional is false of each act when  $o_1$  and  $o_2$  are equally good.<sup>12</sup>

Thus, while Brown's argument is neutral, it does not establish a deep fault line. And for this reason it fails to vindicate traditional wisdom that cases of deontic constraints form a deep fault line.

---

<sup>12</sup> Of course, Brown's argument does not establish the same shallow fault line as cases of deontic options do (Brown's concerns *Right-to-Left*, cases of deontic options concerns *Left-to-Right*). Nonetheless, the fault line it establishes is not deep.

## 1.5 Impossibility

And in fact, the failure of Brown's argument illustrates why the traditional wisdom must be mistaken: it is impossible to show that there is a deep fault line on neutral grounds. Notice that in order to ensure neutrality Brown considered each possible way the consequentialist could rank outcomes. One of the orderings involved having the outcomes being ranked the same.

This ordering of outcomes cannot be ruled out on neutral grounds. After all, there is nothing about the structural or logical properties of goodness that entails that certain outcomes have to be ranked ahead of others. Rather, the only grounds that there might be for insisting in a particular case that some outcome has to be ranked above another would have to at least be minimally non-neutral.

And this shows that the traditional wisdom is false: Ranking outcomes as tied cannot be ruled out on neutral grounds. If we rank outcomes as tied, we will not be able to show that *Left-to-Right* or *Compelling Idea* is incompatible with cases of deontic constraints.

Having seen that the traditional wisdom fails, two natural strategies for establishing the fault line and correctly describing its importance suggest themselves. One strategy is to give up on establishing a deep fault line. And Brown's argument shows that if we do this, we can establish shallow fault line on neutral grounds. Another strategy is to give up on establishing a fault line on neutral grounds. This is the strategy that I will pursue.

## 2. Unanimity and Deontic Constraints Structured as Prisoner's Dilemma

In pursuing this strategy, I will try to take a minimal departure from neutrality in an attempt to stay as close as possible to the roles that the traditional wisdom claimed for the fault line and thereby try to save as much of the standard practice of ethical theorists as possible. The trouble with neutrality, recall, was that it allowed for us to rank the outcomes the same. What I will do is isolate the weakest and most plausible principle that I can think of that will rule out axiologies that rank outcomes the same in cases of deontic constraints. By doing this, I hope to establish a deep fault line on non-neutral but *modest* grounds.

Before I delve into this task, one caveat is in order. I know of now simple and plausible way of determining how modest an axiological assumption is. So in what follows I will present the most modest assumption that I can think of and leave it to the reader (and a few footnotes) to assess whether there may be some even more modest assumption that could do the work that we require.

### 2.1 Unanimity

The principle that I will rely on is the following:

*Unanimity* If every agent ought to prefer some outcome,  $o_i$ , to another outcome  $o_j$ , then  $o_i$  is better than  $o_j$ .

*Unanimity* is a high level claim about the relationship between what agents ought to prefer and what is good. *Unanimity* does not follow from purely structural or logical features of goodness so it is non-neutral. It is however a very weak and highly plausible principle. And as I will argue later in this section, it is a principle that will force us to rank one outcome higher than another in cases of deontic constraints. If we can establish that cases of deontic constraints form a deep fault line on the basis of *Unanimity*, I believe that this will be as close as we can come to vindicating the traditional wisdom.

I will not offer any argument directly in favor of *Unanimity*. This is because any argument that I can think of relies on premises that are jointly less plausible than *Unanimity* itself. Instead, I will try to bring out the inherent plausibility of this principle by clarifying what it says and locating it in the broader context of issues at the intersection of moral philosophy and social choice theory.

Let me begin by clarifying what *Unanimity* says. In particular, I want to discuss the sense of ‘ought’ that figures in this principle. I intend this ‘ought’ to be one that is *objective* and *all-things-considered*. It is *objective* in the sense that it depends on the facts and not on the epistemic state of the agent (e.g. if an agent fails to know that pushing button A will greatly benefit everyone and pushing button B will greatly harm everyone, it is still true in this sense that the agent ought to prefer the outcome of pushing button A to pushing button B). It is *all-things-considered* in the sense that what we ought in this sense to do is determined by all the practically relevant considerations rather than a subset of them (e.g., the moral considerations)

This sense of ‘ought’ is also what we might call the ‘ought’ of fittingness.<sup>13</sup> To see what I mean by this, consider two examples. Suppose that an evil demon threatens to kill me unless I believe that  $2+2=5$  or an evil demon threatens to kill me unless I prefer a world in which I experience total agony to a world in which I do not. There is a sense in which I ought to believe that  $2+2=5$  and ought to prefer the world with total agony—I will stay alive. But there is another perfectly good sense in which I ought not to believe  $2+2=5$  and ought to not prefer the world with total agony. After all, I have plenty of evidence that  $2+2\neq 5$  and there is nothing but agony in that world. The ‘ought’ of fittingness is this later sense of ‘ought’ on which I ought to not have this belief and ought not to have this preference. Very roughly, in the sense of ‘ought’ that I have in mind, whether you ought to prefer some outcome to another one depends on the properties of the outcomes rather than the properties of your having the preference or belief. In the terminology of the philosophical literature of fitting attitudes, the fact that

---

<sup>13</sup> Thanks to Sam Shpall for pressing me to be clearer about this.

you will be killed for refraining from forming a preference or belief is simply the wrong kind of reason to ground the claim that you ought to have this preference or belief in the fittingness sense of ‘ought’.<sup>14</sup>

It is important to be clear about this because certain consequentialist theories purport to tell us which preferences we ought to have.<sup>15</sup> According to these theories, we ought to have a preference when having this preference would lead to the best outcome. These theories are best understood as being about a different sense of ‘ought’ than the one that I have in mind. This is because they blatantly invoke considerations that are the wrong kind of reason to bear on whether I ought to have a preference in my sense of ‘ought’.<sup>16</sup>

This point is noted in different terminology in Derek Parfit’s *Reasons and Persons*. Parfit distinguishes between an agent’s theory-given aims and the aim the theory tells the agent she ought to have. The consequentialist-theory-given aim for each agent is the aim of realizing the best outcome. This does not however mean that consequentialism says that each agent ought to aim to realize the best outcome for it may be that aiming to realize the best outcome does not lead to the best outcome. In Parfit’s terminology, my sense of ‘ought’ is the one where an agent ought to prefer  $o_i$  to  $o_j$  just in case  $o_i$  better fulfils the aims given to that agent by the correct theory.

With this clarification in hand, let me turn to locating *Unanimity* within the context of moral philosophy and social choice theory. Social choice theory is a field where, among other things, theorists are interested in trying to understand whether and to what extent the so-called social good, which we can think of as goodness, depends on the goods of individuals.<sup>17</sup> In that field, a very similar principle has been proposed. Though this principle is sometimes called ‘unanimity’ as well, I will refer to it by another common name in order to avoid confusion:

*Weak Pareto Efficiency*: If an outcome  $o_i$  is better for every agent than an outcome  $o_j$ , then outcome  $o_i$  is better than outcome  $o_j$ .

Both *Weak Pareto Efficiency* and *Unanimity* claim that there is a connection between goodness and some other normative notion. And while there are interesting question about what explains this connection, both principles are appealing in their own right. The main difference between *Weak Pareto Efficiency* and *Unanimity* is that *Weak Pareto Efficiency* concerns a connection between the goods of individuals and goodness while *Unanimity* concerns a connection between what each individual ought to prefer and

---

<sup>14</sup> See D’Arms and Jacobsen 2000 and Rabinowicz and Rønnow-Rasmussen 2004

<sup>15</sup> I have so-called global consequentialism particularly in mind, see Pettit and Smith 2000, Driver 2012: ch. 7, Chappell 2012.

<sup>16</sup> Cf. Way Forthcoming’s criticism of the project of explaining the fittingness sense of ‘ought’ in terms of value.

<sup>17</sup> See Gaertner 2009 for an introduction.

goodness. One way to see why *Unanimity* is overwhelmingly plausible is to see why it is not vulnerable to certain objections that you might give to the already plausible *Weak Pareto Efficiency*.

For example, if we understand the good of an agent as her well-being, we might worry that *Weak Pareto Efficiency* is not sufficiently sensitive to the distribution of well-being. In particular, suppose in  $o_1$  person A's well-being is 5 and person B's well-being is 6 and in  $o_2$  person A's well-being is 6 and person B's well-being is 15. In this case *Weak Pareto Efficiency* would tell us  $o_2$  is better than  $o_1$ . This might seem worrisome on the grounds that  $o_2$  is actually worse than  $o_1$  because there is much greater inequality in  $o_2$ .

While I don't want to endorse this objection to *Weak Pareto Efficiency*, it is important to see that it is much less plausible as an objection to *Unanimity*. This is because it is natural to think that if equality considerations matter for the goodness of outcomes, they also matter for what agents ought to prefer. Thus, if we focus on what each of the agents ought to prefer, it will turn out that each of them ought to prefer  $o_1$  to  $o_2$ . In this way, *Unanimity* is not vulnerable to this objection to *Weak Pareto Efficiency*.

More generally, *Unanimity* is not vulnerable to a whole host of objections to the idea of understanding goodness in terms of the well-being of individuals.<sup>18</sup> This is because whatever other features contribute to goodness, they also plausibly contribute to what individuals ought to prefer. This illustrates why *Unanimity* is more plausible than the already plausible *Weak Pareto Efficiency*.<sup>19</sup>

Having explained why *Unanimity* is plausible, I will now turn to clearing the way for my argument that cases of deontic constraints form a fault line in ethical theory.

## 2.2 Prisoner's Dilemmas

The first thing to notice is that we have some work to do before we can use *Unanimity* to show that cases of deontic constraints form a fault line. This is because we cannot simply apply *Unanimity* to Mary's case in order to get the result that we want. What we are trying to show in Mary's case is that the outcome in which Mary murders is better than the outcome in which Mary does not murder. We could use *Unanimity* to get this result if we could show that everyone ought to prefer the outcome in which Mary murders to the outcome in which she does not murder. Unfortunately, we do not have any reason to think that this must be true. In fact, it seems plausible enough to think that Mary ought to prefer the

---

<sup>18</sup> To give another example, it would avoid the non-identity problem for reducing goodness to well-being that can be found in Parfit 1984: ch. 16.

<sup>19</sup> I have chosen to focus on the philosophical difference between *Unanimity* and *Weak Pareto Efficiency*. I have, because of this, ignored the issues in social choice theory that might bear on the acceptability of *Unanimity*. In particular, it is worth noting that *Weak Pareto Efficiency* is a premise in both Arrow's impossibility result and Sen's liberal paradox (see Arrow 1963, Sen 1971). Unfortunately, I do not have space in this paper to consider these results in detail. It is however worth making two points. First few reject *Weak Pareto Efficiency* on the basis of these results and variants of these results can be established without assuming *Weak Pareto Efficiency* (see Gibbard 1974). Second the other assumptions of these proofs (e.g., unrestricted domain) are highly suspect if we consider them for the notion of ought to prefer that figures in *Unanimity*.

outcome in which she refrains from murdering to the outcome in which she murders. For this reason, *Unanimity* is not of any help “out of the box”.

What I am going to do in this subsection is set things up so that we can apply *Unanimity* to get the result that we want. This will require us to step back from the issues that we have been discussing so far in two respects. First for most of this section, we will set aside issues concerning right and wrong and turn to considering what we have reason to do. As I said before just as common sense moral thought recognizes that Mary’s case is a case where it is wrong to murder even to prevent further murders, common sense practical thought recognizes that Mary has better reason to refrain from murdering than to murder. So for the next two subsections we will focus on whether the consequentialist theory of practical reason that we introduced in the beginning of the paper—the theory according to which an agent has more reason to do *a* than to do *b* just in case the outcome of *a* is better than the outcome of *b*—can accommodate such cases. We will return to discuss right and wrong in §2.4 where I will show how our conclusions about practical reason can be generalized.

Second, we will have to take a moment to discuss the prisoner’s dilemma. This is because I will be arguing that if we set up cases of deontic constraints so that they have the structure of a prisoner’s dilemma, we can use *Unanimity* to show that cases of deontic constraints form a fault line. For those who are familiar with game theory and social choice theory, this result may come as no surprise because prisoner’s dilemmas are cases where agents who act rationally end up realizing an outcome that is Pareto dominated by another outcome. But for those who work in moral philosophy and for those who want to be careful about how we generalize these claims from other fields to issues in moral philosophy, it will be worth taking the time to spell out this argument in detail.

So let’s consider the prisoner’s dilemma as it is classically presented. In the classic story, we have two people who are being questioned by the police about some crime that they are suspected of committing. Each person faces a choice of agreeing to testify against her partner (defect) or remaining silent (cooperate). If one person chooses to testify (i.e., defect) and the other chooses to remain silent (i.e., cooperate), then the defector goes free and the cooperator gets ten years in jail. If both cooperate, they each get one year in jail. If both defect, they each get three years. We can summarize this situation like this:



Table 1

		Person B	
		Cooperate	Defect
Person A	Cooperate	1 year each	A: 10 years, B: 0 years
	Defect	A: 0 years, B: 10 years	3 years each

After introducing cases like this, it is typical in game theory to argue that the rational thing for each agent to do in this case is defect. The argument goes like this: Consider what A should do. If B cooperates, then A should defect because A thereby avoids having to spend a year in jail. And if B defects, A should defect as well because A thereby avoiding having to spend seven additional years in jail. So no matter that B does, the rational thing for A to do is defect. And by analogous reasoning the rational thing for B to do is defect. So the rational thing for both A and B to do is defect no matter what the other one does.

In game theory, this “dominance” argument is taken to establish that the rational thing for each agent to do is defect no matter what the other agent does. There are a number of issues that bear on whether this kind of dominance argument is sound in a given context. And there are still further difficulties for this line of thought because many people do not believe that our common sense moral or practical thinking says that it is rational for both agents to defect in the case that I described above. Luckily, our approach will not be encumbered by these putative difficulties because we will not be using these cases in the exact same way game theorists use them.

Instead, we will adopt a slightly different perspective that is more suited to discussing what we have reason to do. First we will assume that the individuals in the case have no prior relationship with one another, that they have no ability to communicate with one another (in fact, we may even assume they do not know of one another’s existence), and that neither will ever be in this situation again.

Next though the prisoner’s dilemma that I presented concerned the time individuals will spend in jail and appears to assume that all an individual should care about is how much time she spends in jail, we will generalize this dilemma so it does not narrowly focus on jail time. The way to do this is to return to our talk of what individuals ought to prefer in the sense of ‘ought’ identified in the last section. That is, instead of focusing narrowly on the jail time, we can look at how a given outcome would rank according to the preferences each agent ought to have.

According to this perspective, a prisoner’s dilemma has the following structure:

Table 2

		Person B	
		Cooperate	Defect
Person A	Cooperate	2 <sup>nd</sup> best for each	A: Worst B: Best
	Defect	A: Best B: Worst	3 <sup>rd</sup> best for each

where by ‘best’ ‘2<sup>nd</sup> best’ etc, we are referring to the ranking of the outcome according to the preferences each agent ought to have and where ‘defect’ and ‘cooperate’ are mere labels for the acts that have these outcomes (and so are not taken to suggest anything about the possibility of cooperating with others or defecting from agreements).<sup>20</sup> To see that this really is the natural generalization that I just described, we need only check that this is how the outcomes in *Table 1* would rank if each agent’s only aim was to minimize her time in jail.

In cases with this structure, we can draw certain conclusions about what we have reason to do. To see how, recall our discussion of Portmore and Parfit. Parfit observed that we can think of theories of, e.g., practical reason as giving agents aims and the theory given aims of agents are the aims they ought to have in the sense of ‘ought’ that I identified in the last section. Portmore observed—and we have followed him in this—that we can think of the “theory” of common sense moral and practical thought as giving agents different aims and we can think of these aims as determining a ranking of outcomes for each agent.

So if we interpret *Table 2* as describing such a ranking, we can draw the conclusion that each agent has most *objective* practical reason to defect no matter what the other agent does (where again by objective I mean that these reasons depend on the facts and not on the agent’s epistemic state). The argument for this conclusion is this: Suppose B cooperates, then A has most reason to defect because, by the description of the case, A ought to prefer the outcome of defecting to the outcome of cooperating. Suppose B defects, then A has most reason to defect because, by the description of the case A ought to prefer the outcome of defecting to the outcome of cooperating. So no matter what B does A has most reason to defect. A similar argument, of course, works for B.

The final step to take is to construct cases of deontic constraints so that they have the structure of a prisoner’s dilemma. Consider then the following admittedly artificial but nonetheless instructive example. Person A has contracted three assassins to kill three innocent people. And suppose Person B has contracted three different assassins to kill three different innocent people. A and B are now in

---

<sup>20</sup> Cf. Parfit 1984: 91’s “each-we” dilemmas.

different rooms. In these rooms, there is a button and each faces the choice of whether or not to push it. An automatic email will be generated that calls off a certain number of assassins depending on what pattern of buttons are pushed. The table summarizes what happens depending on the combination of buttons pushed where we read ‘Defecting’ as pushing the button and ‘Cooperating’ as not pushing the button:

		Person B	
		Cooperate	Defect
Person A	Cooperate	1 murder each	A: 3 murders, B: 0 murders
	Defect	A: 0 murders, B: 3 murders	2 murders each

Common sense practical thought says that in this case each agent has most reason to defect no matter what the other agent does. To see this, first suppose B were to cooperate. In this situation, A faces a choice of whether to murder in order to prevent two murders by B. This is analogous to Mary’s choice of whether to murder to prevent two murders and common sense practical thought says that Mary has most reason to not murder. Though the case we are considering now has some additional complications, there is no good reason to think that these complications should alter the verdicts of common sense thought about what A has most reason to do. After all, A has no prior relationship with B, has no ability to communicate with or otherwise affect B’s actions (in fact, we may even assume A does not know of B’s existence), and A will never be in this situation with B again. So common sense practical thought says that A has most reason to not murder. Next suppose B were to defect. In this situation, A again faces the choice of whether to murder to prevent B from committing two murders. And again common sense practical thought suggests that A has most reason to not murder. Of course, similar results apply to B. Thus, no matter what the other agent does each agent has most reason to defect.

To put this same point in a way that makes the connection to what the agents ought to prefer more vivid, recall following Portmore that a standard way of understanding cases of deontic constraints is as cases where an agent has special concern for *her murders*. Given this, the ranking of these outcomes according to the preferences A and B ought to have is the one in *Table 2*. Since A has special concern for her murders she will most prefer to defect when B cooperates, next most prefer to cooperate when B cooperates, third most prefer to defect when B defects, and prefer least of all to cooperate when B defects. Analogously, B prefers most to defect when A cooperates, prefers next most to cooperate when A cooperates, prefers third most to defect when A defects, and prefers least of all to cooperate when B

cooperates. For this reason, this is a case of deontic constraints that has the structure of the prisoner's dilemma.

### 2.3 Establishing the Fault Line

I now want to argue that consequentialism cannot get us the result that each agent has most reason to defect no matter what the other agent does. To do this, I will show what the consequentialist would have to say in order to get this result and show that this not only is unintuitive but also violates *Unanimity*.

So, recall, the consequentialist theory of practical reason says that an agent has more reason to do *a* than to do *b* just in case the outcome of *a* is better than the outcome of *b*. Four claims would have to be true for this theory to get the result that each agent has most reason to defect no matter what the other agent does. First, in order to ensure that A has most reason to defect if B cooperates, the consequentialist must claim that the outcome where B cooperates and A defects is better than the outcome in which both A and B cooperate:

		Person B	
		Cooperate	Defect
	Cooperate	∧	
Person A			
	Defect		

Second, in order to ensure that B has most reason to defect if A cooperates, she must claim that the outcome where A cooperates and B defects is better than the outcome in which both A and B cooperate:

		Person B	
		Cooperate	Defect
	Cooperate		<
Person A		∧	
	Defect		

Third, in order to ensure that A has most reason to defect if B defects, the consequentialist must claim that the outcome in which both A and B defect is better than the outcome in which B defects and A cooperates:

		Person B	
		Cooperate	Defect
	Cooperate	<	
Person A		^	^
	Defect		

Fourth, in order to ensure that B has most reason to defect if A defects, she must claim that the outcome in which both A and B defect is better than the outcome in which A defects and B cooperates:

		Person B	
		Cooperate	Defect
	Cooperate	<	
Person A		^	^
	Defect	<	

Summarizing then, in order to get the result that each agent has most reason to defect no matter what the other agent does, the consequentialist must claim:

Table 3

		Person B	
		Cooperate	Defect
	Cooperate	Worst	2 <sup>nd</sup> or 3 <sup>rd</sup> Best
Person A			
	Defect	2 <sup>nd</sup> or 3 <sup>rd</sup> Best	Best

So *Table 3* gives us the ranking of outcomes that the consequentialist would need to have in order to be compatible with cases of deontic constraints that have the structure of a prisoner's dilemma.

But this ranking of outcomes violates *Unanimity*. To see this, recall that *Table 2* shows us what the prisoner's dilemma looks like in terms of what agents ought to prefer:

Table 2

		Person B	
		Cooperate	Defect
	Cooperate	2 <sup>nd</sup> best for each	A: Worst B: Best
Person A			
	Defect	A: Best B: Worst	3 <sup>rd</sup> best for each

The first thing to notice is that comparing *Table 3* and *Table 2* tells us that if the consequentialist is to get the result that each agent has most reason to defect, she must claim that the outcome that each agent ought to rank third is best and that the outcome that each agent ought to rank second is worst. This is incompatible with *Unanimity*. As *Table 2* shows, every agent ought to prefer the outcome in which both cooperate to the outcome in which both defect. Recall now that *Unanimity* says that if every agent ought to prefer  $o_i$  to  $o_j$  then  $o_i$  is better than  $o_j$ . So it follows from this and *Table 2* that the outcome in which both cooperate is better than the outcome in which both defect. But *Table 3* contradicts this claim because it says the outcome in which both cooperate is worst and the outcome in which both defect is best.

In other words, we have successfully argued from *Unanimity* that consequentialism is incompatible with cases of deontic constraints. We can illustrate this same argument as it applies to the more concrete case of the prisoner's dilemma involving murders:

		Person B	
		Cooperate	Defect
	Cooperate	1 murder each	A: 3 murders, B: 0 murders
Person A			
	Defect	A: 0 murders, B: 3 murders	2 murders each

In order to get the result that each agent has most reason to defect no matter what the other agent does in this case, the consequentialist must say the best outcome is the one in which each murders two and the worst outcome is the one in which each murders one. But *Unanimity* entails that this cannot be right.<sup>21</sup>

---

<sup>21</sup> Note that this case shows that everyone including people other than A and B prefer the outcome in which both cooperate to the outcome in which both defect. (At least this is so if we think of the same two people dying in the case of both cooperating as die in the case in which both defect).

Thus, while our argument is not neutral, it is based on the very weak and plausible principle *Unanimity*. Since *Unanimity* forces us to rank one outcome ahead of another, it avoids the obstacle that Brown's argument faced. It was because Brown's argument did not force us to rank one outcome ahead of another that it was unable to establish a deep fault line.<sup>22</sup> For this reason, I conclude that though the traditional wisdom is false, we *can* show on modest ground that there is a deep fault line in ethical theory and thereby come as close as, I believe, we can to vindicating the traditional wisdom and the practices of ethical theorists.

## 2.4 Morality

This result is enough on its own to establish my thesis. This thesis concerned what I called the consequentialist program of vindicating the idea that goodness or value is central. That program was committed to explaining the notions of right and wrong and what we have reason to do in consequentialist terms. I have shown on modest ground that this program is deeply incompatible with cases of deontic constraints.

But in showing this, I have focused exclusively the consequentialist theory of reasons. So it is worth taking a moment to close this section by providing a simple generalization of my argument so that it applies to the consequentialist moral theory. To see how this generalization works, return to the case prisoner's dilemma involving murders that I described above. Common sense moral thinking tells us that in this case it would be wrong to cooperate and right to defect no matter what the other agent does. Since we have already seen that *Unanimity* rules out the ordering of outcomes needed to get this result, this suffices to show that at least this case is incompatible with consequentialism. Similar results apply to other cases structured like prisoner's dilemmas where our judgments about what is right and wrong and what we have most reason to do line up.<sup>23,24</sup>

---

<sup>22</sup> It is straightforward to directly verify (though perhaps not immediately obvious) that *Left-to-Right*, *Right-to-Left*, and *Compelling Idea* must fail in this case for this reason.

<sup>23</sup> We might worry that this way of generalizing things does not show that cases of moral deontic constraints are incompatible *as such* with consequentialism. So we might hope for a general recipe for creating cases of deontic constraints that have the structure of a prisoner's dilemma that does not require us to appeal to judgments about practical reason as well as judgments about morality. To do this, we would have to make one of two modifications. We could keep *Unanimity* as it stands and claim that our moral-theory-given aims determine not what we morally ought to prefer but what we ought to prefer all-things-considered. If we pursue this route, we must either presuppose some sort of moral rationalism or say what we have reason to do and what we ought to prefer are connected in a more complicated way than I have assumed. Alternatively, we may adopt *Unanimity\** which says that if everyone morally ought to prefer  $o_1$  to  $o_2$ , then  $o_1$  is better than  $o_2$  and stick with the idea that our moral-theory-given aims concern what we morally ought to prefer. Earlier we, of course, also accepted *Unanimity*. So if we take this route, we will either need to abandon *Unanimity* or adopt a form of moral rationalism that will ensure that *Unanimity* and *Unanimity\** never conflict. I will not delve into which of these routes is preferable. I will rest content with having shown that cases of deontic constraints in which practical reason and morality align are incompatible with consequentialist theories of practical reason and morality.

<sup>24</sup> Parfit 1984: 103-108 argues that no moral theory should say that each individual is morally required to defect in the prisoner's dilemma. This argument is no threat to my generalization. Parfit's argument is intended to show that moral theories that recognize cases of deontic constraints are false. I am remaining agnostic on whether a moral theory should recognize cases

### 3. Oddie and Milne

Having established this, I want to close the paper by showing how this argument allows us to diagnose where Oddie and Milne's argument in "Act and Value" goes wrong.

The main goal of their paper is to establish that a consequentialist moral theory can "represent" any set of deontic verdicts. Oddie and Milne use the term 'consequentialist moral theory' slightly differently than I do. According to my usage each act is associated with a unique outcome and whether an act is wrong is determined by the value of that outcome. Oddie and Milne however associate each act with a set of outcomes together with a probability distribution over those outcomes. Their idea, then, is that an agent is required to perform an action just in case the expected value of that act is greater than that of any alternative where the expected value of an act is understood as the sum of the value of its outcomes weighted by the probability of their occurring. And their claim is that the consequentialist can get the correct results in cases of deontic constraints simply because we can prove that there are always claims about value that the consequentialist can make that will ensure that they capture any set of verdicts about what is right and wrong.

Oddie and Milne's proof rests on relatively conservative assumptions. In particular, they assume that there is a finite set of agents, a finite set of times, and a (possible infinite) set of worlds. They assume that an agent at a world and at a time has a finite set of acts she could perform and each act has a finite set of outcomes. They assume that for each world and time there is a unique distribution of probabilities over outcomes. Finally, they assume that the *better than* ordering on outcomes has certain minimal structural properties (transitivity etc.). They then show that given any claims about which acts for which agent are required, permitted, etc. at a world and time, there is some ordering of outcomes according to how valuable they are that, when combined with consequentialism, yields these results about what is required, permitted, etc. at that world and time. And more generally, it seems that this proof will show that given a finite set of acts ordered according to how much reason there is, there is always a way of ranking outcomes according to how good they are that can get you that result. This is the highly general result that got us worried in the first place about whether the traditional wisdom was true. And it looks like this same result should get us worried about our argument as well.

Fortunately, the work that we have done in this paper will enable us to show that Oddie and Milne's result is no threat to our argument. The reason why it is no threat is what might at first seem to be a strength of their proof—the conservative assumptions that I just outlined. The trouble is that by only making conservative assumptions about goodness, Oddie and Milne's result fails to establish that

---

of deontic constraints in this paper. My concern only has been to show on modest grounds that the consequentialist program is incompatible with cases of deontic constraints.



there is an ordering on outcomes that is actually the *goodness* ordering that we can use to get the deontic verdicts we want.

Here is one stark way of seeing this. Suppose for example I have an *anti-consequentialist theory of reasons* that says that an agent has better reason to do *a* than to do *b* just in case the expected value of doing *a* is worse than the expected value of doing *b*. Oddie and Milne's proof makes such conservative assumptions about the nature of the ordering of outcomes that the *worse than* relation will count as an ordering on outcomes that satisfies these assumptions. And their proof will show, in effect, that there is always some way of ordering outcomes that makes it so the anti-consequentialist theory of reasons gets the right deontic verdicts.

But this result does not lend much credibility to the anti-consequentialist theory of reasons. After all, the following example still suffices to show that the anti-consequentialist theory of reasons is false: You have an option of punching John in the face or not. Presumably you have more reason to refrain from punching John in the face than to punch him. But the outcome in which you punch John in the face is clearly worse than the outcome in which you do not punch him the face. So the anti-consequentialist theory entails that you do not have most reason to refrain from punching John.

This case is a counterexample despite Oddie and Milne's result because Oddie and Milne's result only tells us which claims about value *would have to be true* if the anti-consequentialist theory were to get the right result in this case. It does not tell you anything about whether those claims are actually true. And in this case, in order for the anti-consequentialist theory to get the right result it would have to be true that the outcome in which you refrain from punching John is worse than the outcome in which you punch John. But since we know that this claim is false, Oddie and Milne's result is of no help.

For similar reasons, Oddie and Milne's result does not show that consequentialist theories get the result that each agent has most reason to defect no matter what the other agent does. For the consequentialist to get this result, we saw they must say the outcome in which both defect is best and the outcome in which both cooperate is worse. But this ranking of outcomes is not a goodness ranking. This is because it violates *Unanimity*.<sup>25</sup>

---

<sup>25</sup> Interestingly, Oddie and Milne actually recognize that their result tells us roughly this about the prisoner's dilemma. They say it tells us what would have to be true is that an outcome is best if it contains "agents' pursuing their own good irrespective of the goodness of the outcome" (74). And they quite rightly report that this claim about value "sacrifices some clear axiological intuitions". And so they go onto reject it.

Surprisingly, Oddie and Milne seem to take this as some evidence *for* consequentialism. While I am not sure myself why they think this, I suspect that it is due to factors that are unique to their discussion. In particular, they are thinking of the prisoner's dilemma in purely self-interested terms rather than more generally as arising from the fact about what agents ought to prefer. In effect, they see this result as an argument against an egoist moral theory. But we are using prisoner's dilemmas in a different way. In our context, we are discussing cases that have the structure of a prisoner's dilemma but do not arise because of pure self-interest. We are also discussing this issue from the perspective of practical reason as well as morality. Because of these differences, we should not follow Oddie and Milne in thinking that this is evidence for consequentialism

So what looked like a virtue of Oddie and Milne’s proof—that it relied on very few assumptions—turns out to be a vice. The proof does not assume enough to ensure that the ordering of outcomes it generates in order to get the right results in the prisoner’s dilemma is actually an ordering according to the *goodness* of those outcomes. And the very same facts that we identified earlier suffice to show that the ordering that the proof generates cannot be the goodness ordering. It cannot be the goodness ordering because the goodness ordering obeys *Unanimity*.

This nicely corroborates our diagnosis of the status of the traditional wisdom. That wisdom, I argued, is false because we cannot show deontic constraints are incompatible with all three these entailed by consequentialism that we identified earlier without making some non-neutral axiological assumptions. Since Oddie and Milne’s assumptions are all neutral assumptions, they, in effect, provide a formal proof that tells us exactly how to rank outcomes in order for consequentialism to allow for cases of deontic constraints.<sup>26</sup> But once we allow some minimal non-neutral axiological assumption, we are in a position to see that their proof is no threat to the idea that cases of deontic constraints form a fault line.

#### 4. Conclusion

In sum, our argument that cases of deontic constraints form a deep fault line gives us the resource to reject on modest grounds the results from Oddie and Milne that got us worried in the first place. Of course, none of this shows that consequentialism is false or that there are no cases of deontic constraints. Consequentialists are free to argue—as they have—that common sense moral and practical thought’s recognition of cases of deontic constraints is a mistake. And those who are convinced that these judgments are correct can rest content that it follows from them that consequentialism is false.

That said, the work that we have done here shows that consequentialists and their critics should turn their focus away from simple cases of deontic constraints and toward *Unanimity* and cases of deontic constraints that have the structure of the prisoner’s dilemma. And it tells us that both the traditional wisdom and Oddie and Milne are mistaken. The traditional wisdom is mistaken in thinking that a deep fault line can be established on neutral grounds; Oddie and Milne’s in thinking that it cannot be established at all. The truth is that we need only make a modest axiological assumption in order to vindicate the important thought that cases of deontic constraints form a deep fault line in ethical theory.<sup>27</sup>

---

<sup>26</sup> Oddie and Milne even deny that a shallow fault line can be established on neutral grounds. I do not find their argument persuasive. It requires the premise that cases like Brown’s are not cases where an agent can genuinely act. Since I find this premise implausible, I have not relied on this response to Brown (see Carlson 1995: ch. 3 for more on this issue).

<sup>27</sup> I would like to thank the USC dissertation group, the audience at Texas Tech, Howard Curzer, Steve Finlay, Alex Grzankowski, Chris Hom, Ben Lennertz, Ryan Millsap, Danny Nathan, Doug Portmore, Walt Schaller, Sam Shpall, Justin Snedegar, Aness Webster, and Ralph Wedgwood for comments. Thanks most of all to Mark Schroeder for advice and criticism

## References

- Arrow, Kenneth. 1963. *Social Choice and Individual Values* 2<sup>nd</sup> Eds. New Haven: Cowles Foundation for Research in Economics at Yale University.
- Bennett, Jonathan. 1989. "Two Departures from Consequentialism" in *Ethics* 100: 54-66.
- Brook, Richard. 1991. "Agency and Morality" in *The Journal of Philosophy* 88: 190-212.
- Brown, Campbell. 2011. "Consequentialize This" in *Ethics* 121: 749-771.
- Carlson, Erik. 1995. *Consequentialism Reconsidered*. Dordrecht: Kluwer.
- Chappell, Richard. 2012. "Fittingness" in *The Philosophical Quarterly* 62: 684-704.
- D'Arms, Justin and Daniel Jacobson. 2000. "Sentiment and Value" in *Ethics* 110: 722-748
- Darwall, Stephen. 1983. *Impartial Reason*. Ithaca: Cornell University Press.
- . 1986. "Agent-Centered Restrictions from the Inside Out" in *Philosophical Studies* 50: 291-319.
- Dreier, James. 1993. "Structures of Normative Theories" in *The Monist* 76: 22-40.
- . 2011. "In Defense of Consequentializing" in *Oxford Studies in Normative Ethics* 1: 97-199.
- Driver, Julia. 2012. *Consequentialism*. London: Routledge
- Foot, Philippa. 1985. "Utilitarianism and the Virtues" in *Mind* 94: 196-209.
- Gaertner, Wulf. 2009. *A Primer in Social Choice Theory* Revised Eds. Oxford: Oxford University Press.
- Gibbard, Allan. 1974. "A Pareto-Consistent Libertarian Claim" in *Journal of Economic Theory* 7: 388-410.
- Hare, R.M. 1981. *Moral Thinking*. Oxford: Oxford University Press.
- Hurley, Paul. 1997. "Agent-Centered Restrictions" in *Ethics* 108: 120-146.
- Kagan, Shelly. 2001. "Thinking about Cases" in *Social Philosophy & Policy* 18: 44-63.
- Kamm, F.M. 2008. *Morality, Mortality*, vol 2. Oxford: Oxford University Press.
- List, Christian. 2006. "Republic Freedom and the Rule of Law" in *Politics, Philosophy, and Economics* 5: 201-220.
- Louise, Jennie. 2004. "Relativity of Value and the Consequentialist Umbrella" in *The Philosophical Quarterly* 54: 518-536.
- Moore, G.E. 1903. *Principia Ethica*.
- Nagel, Thomas. 1978. *The Possibility of Altruism*. Princeton: Princeton University Press.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Oddie, Graham and Peter Milne. 1991. "Act and Value" *Theoria* 57: 42-76.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pettit, Philip. 2011. "The Instability of Freedom as Noninterference" in *Ethics* 121: 693-716.
- Pettit, Philip and Geoffrey Brennan. 1986. "Restrictive Consequentialism" in *Australasian Journal of Philosophy* 64: 438-455.
- and Michael Smith. 2000. "Global Consequentialism" in Brad Hooker, Elinor Mason, and Dale Miller ed. *Morality, Rules, and Consequences*. Edinburgh: Edinburgh University Press. 121- 133.
- Portmore, Douglas. 2005. "Combining Teleological Ethics with Evaluator Relativism" in *Pacific Philosophical Quarterly* 86: 95-113.
- . 2009. "Consequentializing" in *Philosophy Compass* 4: 329-347.
- . 2011. *Commonsense Consequentialism*. Oxford: Oxford University Press.
- Rabinowicz, Wlodek and Toni Rønnow-Rasmussen. 2004. "Strike of the Demon" in *Ethics* 114: 391-423.
- Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality" in *Philosophy and Public Affairs* 13: 134-171.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.
- Regan, Donald. "Against Evaluator Relativity" in *Philosophy and Public Affairs* 12: 93-112.
- Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge: Belknap Press.
- Scheffler, Samuel. 2003. *The Rejection of Consequentialism* Revised Eds. Oxford: Oxford University Press.
- Schroeder, Mark. 2006. "Not So Promising After All" in *Pacific Philosophical Quarterly* 87: 348-356.
- . 2007. "Teleology, Agent-Relative Value, and 'Good'" in *Ethics* 117: 265-295
- Sen, Amartya. 1971. "The Impossibility of a Paretian Liberal" in *Journal of Political Economy* 78: 152-157.
- . 1982. "Rights and Agency" in *Philosophy and Public Affairs* 11: 3-39.
- Singer, Peter. 1974. "Sidgwick and Reflective Equilibrium" in *The Monist* 58: 490-517.
- . 2005. "Ethics and Intuition" in *The Journal of Ethics* 9: 331-352.
- Smart, J.J.C and Bernard Williams. 1973. *Utilitarianism For & Against*. New York: Cambridge University Press.
- Smith, Michael. 2003. "Neutral and Relative Value after Moore" in *Ethics* 113: 576-598
- . 2009. "Two Kinds of Consequentialism" in *Philosophical Issues* 19: 257-272.
- Way, Jonathan. Forthcoming. "Value and Reasons to Favor" in *Oxford Studies in Metaethics* 8.

---

on every issue at every stage of this project. Finally, I thank the USC Provost's PhD Fellowship, the Flewelling award, and the Russell Fellowship for support.