

Agent-neutral deontology

Author(s): Tom Dougherty

Source: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 163, No. 2 (March 2013), pp. 527-537

Published by: Springer

Stable URL: <http://www.jstor.org/stable/41932683>

Accessed: 03-04-2018 22:52 UTC

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

*Springer* is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*

# Agent-neutral deontology

Tom Dougherty

Published online: 30 October 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** According to the “Textbook View,” there is an extensional dispute between consequentialists and deontologists, in virtue of the fact that only the latter defend “agent-relative” principles—principles that require an agent to have a special concern with making sure that *she* does not perform certain types of action. I argue that, contra the Textbook View, there are agent-neutral versions of deontology. I also argue that there need be no extensional disagreement between the deontologist and consequentialist, as characterized by the Textbook View.

**Keywords** Consequentialism · Deontology · Agent-relative · Agent-neutral · Constraint · Axiology

## 1 The textbook view

A well-known view of the debate between consequentialists and deontologists goes as follows. Consequentialists claim you ought to bring about the best outcome. By contrast, deontologists claim that sometimes you must avoid performing actions of a certain type, even when doing so would lead to otherwise desirable consequences. As a result, deontologists disagree with consequentialists about what you ought to do in certain cases. For example, if a deontologist believes that killing is a type of action that is to be avoided, then she will disagree with a consequentialist about a case like this:

---

T. Dougherty (✉)  
Center for Ethics in Society, Stanford University, 482 Galvez Mall, Stanford, CA 94305, USA  
e-mail: tom.dou@gmail.com

MAFIA. The mafia are credibly threatening to kill two strangers, unless you kill a third.<sup>1</sup>

A consequentialist will say that someone's death is a bad thing, and she may even say that a killing is a particularly bad thing.<sup>2</sup> But her view is that you ought to minimize the number of bad things in the world. You would do this by killing the third stranger. Therefore, the consequentialist will say that you ought to kill the third stranger in MAFIA. Meanwhile, our deontologist will say that it is wrong for you to kill in MAFIA because killing is a type of action to be avoided. The desirable outcome of preventing two other killings is not sufficient reason to perform this type of action. These theories disagree about which actions you ought to perform; as I will put it, they "disagree extensionally," meaning that they disagree about which actions are right and wrong. In light of this extensional disagreement between the consequentialist and deontologist, we arrive at the following claim:

EXTENSIONAL DISAGREEMENT. Consequentialists and deontologists disagree extensionally. Specifically, they disagree about cases like MAFIA.

That is the first plank of what I will call the "Textbook View."

The view continues by aiming to characterize consequentialism and deontology more precisely. It looks for an *intensional* difference between the theories that leads them to this *extensional* disagreement. The Textbook View claims that consequentialists aim at bringing about the outcome that is best from an impersonal point of view. Bringing about the impersonally best outcome is said to be an "agent-neutral" goal: it is the same goal for all agents, whoever they are. Meanwhile, deontologists hold that there are "constraints" (or "restrictions") on certain action types, such as a constraint on killing.<sup>3</sup> A constraint against killing gives an agent a concern with

<sup>1</sup> This case is deliberately schematic in order to guide our focus to the number of possible deaths and the ways that these may come about. So in discussing this case, and all those that follow, let us assume that "all else is equal." (It does not turn out that one of the five is on the brink of discovering the cure for cancer, and so on.)

<sup>2</sup> By saying that killings are worse than deaths, she may claim that you may not kill in the well-known case:

TRANSPLANT. You can kill an innocent person and transplant her organs to save five others.

<sup>3</sup> "A constraint on *what*?", we might ask. Shelly Kagan portrays a constraint as a constraint on maximizing the good. Kagan (1989) reason is that deontologists recognize that there is a "*pro tanto* reason to promote the good" since they think you have a reason to help others. This may make it seem as if a deontologist starts with a consequentialist view, and then adds extras. This portrayal leaves them vulnerable to the charge that their additions are ad hoc, and posited only to save intuitions about cases. But deontologists need not think of their theories in this way. Indeed some, such as Judith Jarvis Thomson, explicitly deny that there is a property of goodness *simpliciter*. They ground an obligation to help others in other ways, such as in a virtue of benevolence (Thomson 1993). Still, the language of "constraints" can survive the observation that deontologists need not be fashioned in the image of consequentialists. Rather than positing "constraints on promoting the good," deontologists could be seen as positing "constraints on action-types." I will use the term "constraint" in this way to denote a deontic prohibition on certain action-types like killing. By contrast, a theory that simply claimed that there are morally relevant thresholds would not thereby count as positing a constraint in this sense. For example, consider the theory that an action is right just in case it produces more happiness than all other actions, unless it produces more than a certain threshold amount of happiness. This threshold theory would not count as positing a constraint in the sense of the word I intend. Thanks to Wayne Davis for prompting me to clarify this latter point.

making sure that *she, the agent herself*, does not kill. Such a constraint is said to be “agent-relative”—the special concern with the agent’s killing is relative to the agent herself.<sup>4</sup> The agent-relativity of the principle shows up in the fact that, in stating the principle, one would need to emphasize the fact that it is the agent’s own actions that are especially prohibited. The Textbook View fixes on the agent-relative/agent-neutral distinction to characterize the intensional difference between consequentialism and deontology. Because consequentialists only care about agent-neutral goals, consequentialism is an agent-neutral theory. Because deontologists claim that agent-relative constraints matter too, deontology is an agent-relative theory:

INTENSIONAL DISAGREEMENT. Consequentialism is an agent-neutral theory.  
Deontology is an agent-relative theory.

This is the second plank of the view. There are no more planks to come. We have met the Textbook View in its entirety. Twenty years ago, Piers Rawling and David McNaughton said it was “fast becoming the standard method of drawing the distinction between consequentialism and deontology.”<sup>5</sup> It remains popular today.<sup>6</sup>

There has been an important challenge to this view. James Dreier has claimed that some consequentialist theories are agent-relative.<sup>7</sup> Dreier conceives of consequentialism as follows: any view that explains what you ought to do in terms of the consequences of your action is consequentialist, even if the view evaluates consequences on an agent-relative basis. For example, on his view, egoism is an agent-relative consequentialist theory because it enjoins an agent to bring about the *consequences* that are best for *the agent herself*.<sup>8</sup> Some people have found Dreier’s conception of consequentialism attractive, and adopted it. But others have not—Daniel Jacobsen, for example, has recently claimed that Dreier’s “will surely remain an idiosyncratic usage.”<sup>9</sup> Their alternative conception of consequentialism is that of Elizabeth Anscombe, who introduced the term to the literature in order to pick out agent-neutral theories like utilitarianism.<sup>10</sup>

In any event, my aim in this paper is to offer a criticism of the Textbook View that is independent of, and potentially complementary to, Dreier’s criticism. My aim

<sup>4</sup> The constraint would have to be time-relative as well as agent-relative if the constraint were to forbid an agent from killing to prevent the agent herself killing at other times. For discussion of time-relativity, see Smith (2009).

<sup>5</sup> McNaughton and Rawling (1992).

<sup>6</sup> See, for example, Kagan (1989), Scheffler (1994) and Pettit (2000). The claim that deontological theories are agent-relative (which is a key claim I argue against in this essay) is even more widespread. As Michael Ridge notes in his helpful discussion of agent-relative and agent-neutral reasons, “the characterization of deontological restrictions as agent-relative (or agent-centered) is close to being an orthodoxy.” Ridge (2008).

<sup>7</sup> Precursors to Dreier’s proposal can be found in Nozick (1974), Sen (1982) and Broome (1991).

<sup>8</sup> Dreier (1993). For more on agent-relative consequentialism, see Portmore (2005); Schroeder (2007); Smith (2009).

<sup>9</sup> Jacobsen (2008, p. 165). Opponents of Dreier’s conception can still accept, as Jacobsen does, the most important point of Dreier’s, which was that some agent-relative theories have a teleological structure that gives a fundamental explanatory role to claims about consequences.

<sup>10</sup> Anscombe (1958).

is show that the Textbook View will not do, by advancing two novel objections to it. Against its second plank, INTENSIONAL DISAGREEMENT, I will argue for the following claim:

AGENT-NEUTRAL DEONTOLOGY. Some deontological theories are agent-neutral.

Then, against the first plank, EXTENSIONAL DISAGREEMENT, I will argue for the following claim:

EXTENSIONAL AGREEMENT. Some theories that the Textbook View would consider consequentialist agree extensionally with deontological theories. Specifically, they agree that you ought not kill in MAFIA.

I will end by considering an alternative to the Textbook View.

## 2 Agent-neutral deontology

I will start by considering a deontological view of killing. According to the Textbook View, if the view holds that there is a constraint against killing, then it would be positing an agent-relative requirement. Following David McNaughton and Piers Rawling, I will state these requirements in terms of rules governing what an agent must ensure comes about. Thus, the Textbook View will characterize the deontological view as positing the following requirement:

AGENT-RELATIVE KILLING CONSTRAINT. Each agent should ensure that *she* does not kill to prevent more killings by others.<sup>11</sup>

We can tell that it is an agent-relative requirement because we need to use the italicized term, “*she*”, in specifying the content of the requirement.<sup>12</sup>

However, I hope that with a little reflection some deontologists would reject the claim that they are concerned with agent-relativity here. Instead, they could simply say that *everyone*, including you the agent, ought to be opposed to your killing the one to save the two. Suppose that there is a bystander watching this grizzly scene. What attitudes should the bystander have towards your choice? What should the bystander want you to do? If the bystander were able to intervene to stop you from killing, ought she do so?<sup>13</sup> A deontologist is free to say that the bystander should be opposed to your killing the single person, even though she knows that this will lead to more deaths overall. So if the bystander were able to intervene to prevent your

<sup>11</sup> Thanks to Wayne Davis for suggesting this formulation of the principle.

<sup>12</sup> In McNaughton and Rawling’s terminology, where “S” stands for “should (ceteris paribus) ensure that” and where square brackets mark out the content of a rule, the logical form of this rule is

(x) [x S (x does not kill to prevent more killings by others)]

Since there is an occurrence of “x” within the square brackets that is bound by the initial universal quantifier “(x)”, this is an agent-relative rule. This formalization aims to capture the intuitive idea that a deontic constraint against killing enjoins an agent to take special care that *she* not kill. McNaughton and Rawling (1991).

<sup>13</sup> Thanks to Caspar Hare for posing these questions.

killing the person you could kill, then she ought to do so. Similarly, the deontologist can say that the bystander ought to prefer that you do not kill. Indeed, I suggest that these are rather attractive claims for the deontologist to make.

What could unify the claim that you the agent should refrain from killing, and the claim that the bystander should be opposed to your killing? One proposal is the following. Both you, the bystander, and everyone else should have the following aim:

AGENT-NEUTRAL KILLING CONSTRAINT. Each agent should ensure that no one kills to prevent more killings by others.<sup>14</sup>

We can see that, as its name suggests, AGENT-NEUTRAL KILLING CONSTRAINT is an agent-neutral principle since it does not give the agent any special concern with her own killings.<sup>15</sup> By positing principles like this, the deontologist could require everyone to “share a moral vision.” That is, she can require everyone to oppose or endorse certain actions, regardless of who they are. Everyone would be united in their view of a particular action, regardless of whether they are the author of the action or not. This is what makes her theory agent-neutral, even though it includes constraints. Thus, a theory that posits constraints like this would be a version of agent-neutral deontology:

AGENT-NEUTRAL DEONTOLOGY. Some deontological theories are agent-neutral.

This, of course, is at odds with how the Textbook View would characterize any theory with constraints. As we saw, according to the Textbook View, the relevant constraint would be an agent-relative constraint. So the Textbook View would get things wrong.

Since the Textbook View purports to taxonomize adequately regions of logical space, it would be interesting enough to note that it fails in this task. But I think that this failure is particularly significant in virtue of the fact that the agent-neutral region is an inhabitable, if not attractive, place for the deontologist to live. This is because there is something appealing about a theory that requires us to share a moral vision. In doing so, we would form a unified moral community, in which we all have the same goals and the same preferences as each other about certain courses of behavior. Why should deontologists have to part with this ideal merely because they hold that some action types have intrinsic wrong-making features? In addition, an agent-neutral version of deontology can avoid a common objection. If a deontological theory posits agent-relative constraints, then it seems rather like it requires an agent to have a concern with avoiding “dirty hands.” It seems that the agent’s only

<sup>14</sup> In McNaughton and Rawling’s terminology, this rule is  
(x)  $\times$  S [(y) y does not kill to prevent more killings by others]

Since there is no occurrence of “x” within the square brackets bound by the initial universal quantifier “(x)”, this is an agent-neutral rule. This is a formalization of the intuitive idea that the content of an agent-neutral requirement does not give the agent any particular concern with *her own* killings (McNaughton and Rawling 1991).

<sup>15</sup> We should not be misled into thinking that this is an agent-relative requirement simply because in some situations it requires the agent not to kill. Here she would simply be “ensuring... that something is true of *her*—but only insofar as she is *one amongst many*.” (McNaughton and Rawling 1991, p. 179).

objection to minimizing the number of killings is that it would be *her* dirtying her hands with the business of killing. It is hard to believe that this could be an important moral reason, let alone one that is so important that it requires increasing the number of people who die. However, it is clear that an agent-neutral version of deontology is immune to this criticism. Since everyone is opposed to the agent's killing, the agent's own opposition is not based on a desire to avoid dirtying her hands.

Why have defenders of the Textbook View overlooked the possibility of such an agent-neutral deontological theory? I suspect that they may have thought that a theory composed of principles like AGENT-NEUTRAL KILLING CONSTRAINT would be too narrow to cover the full range of cases in which a deontologist will judge that an agent ought not kill. They might have thought that if we dreamed up ever more complex cases in which some killings depend on others not occurring, then an agent-neutral theory will not be able to offer the verdicts about these cases that a deontologist desires. But this is false. According to the Textbook View, a deontologist's opposition to killing can be covered by a set of agent-relative principles. Now observe that for every agent-relative principle with this structure,

AGENT-RELATIVE GENERAL CONSTRAINT. Each agent should ensure that she does not  $\Phi$  to prevent a group of others from  $\Omega$ -ing.

there is an agent-neutral principle with this structure,

AGENT-NEUTRAL GENERAL CONSTRAINT. Each agent should ensure that no one  $\Phi$ s to prevent a group of others from  $\Omega$ -ing.

This means that for any deontological theory fashioned out of agent-relative constraints, there is a parallel deontological theory fashioned out of agent-neutral constraints. An agent-neutral version of deontology can be just as broad in scope as its agent-relative cousin.

### 3 Consequentialists against killing

I consider the theory which we just saw to be an agent-neutral version of *deontology* because it posits constraints. That is, it claims that there are some action-types, such as killing, which are forbidden, even if they bring about desirable consequences. I assume that a theory that posits constraints must count as a deontological theory on any acceptable view of the theoretical terrain. I suggest that the theory is not a consequentialist theory because it can be formulated and motivated without appealing to the notion of a good outcome. For example, it could be offered as a version of Kantianism. I will shortly return to the issue of why I suggest this, but first let us consider whether a friend of the Textbook View may try to resist these assumptions. Could she save the Textbook View by resisting this thought, and hold that because it is an agent-neutral theory it must therefore be a consequentialist theory?

I am wary of entering into a merely terminological dispute here. It is open to anyone to stipulate that the term "deontology" is defined in such a way that the term only picks out theories that are agent-relative, and "consequentialist" is defined only to pick out agent-neutral theories. If someone defined the terms in this way,

then she would consider the preceding theory consequentialist rather than deontological. I think that this definition of “deontology” parts too far from how the term is used in moral philosophy for it to be a helpful definition. But more importantly for our purposes, this definition is unacceptable to proponents of the Textbook View. This is because they aim to characterize the dispute between consequentialists and deontologists partly as follows:

EXTENSIONAL DISAGREEMENT. Consequentialists and deontologists disagree extensionally. Specifically, they disagree about cases like MAFIA.

In order to characterize this extensional disagreement, it is important for the Textbook View not to consider a theory that prohibits killing in MAFIA as a consequentialist theory. But a theory with the principle, AGENT-NEUTRAL KILLING CONSTRAINT, does hold that you ought not kill in MAFIA. Therefore, it is important for the Textbook View not to consider such a theory consequentialist.

Indeed, because the Textbook View has this extensional commitment, there is further trouble for it. This is because there is a theory that holds that you ought not kill in MAFIA and the Textbook View would consider consequentialist. According to the Textbook View, consequentialists hold that you ought to bring about the impersonally best outcome. Now consider the following axiological claim:

KILLINGS. If killing  $k_1$  prevents killings  $k_2$  and  $k_3$ , then an outcome with  $k_1$  is *pro tanto* impersonally worse than an outcome with  $k_2$  and  $k_3$ .<sup>16</sup>

According to this theory the causal relations between various killings determines their ultimate values.<sup>17</sup> Let us consider a theorist who holds that you ought to bring about the outcome that is best from an impersonal point of view. Her theory would be an agent-neutral theory and hence one that the Textbook View would consider consequentialist. If she endorsed the axiological claim, KILLINGS, then she would hold that you would bring about an impersonally better outcome in MAFIA if you refrained from killing. Since she holds that you ought to bring about the impersonally best outcome, she would conclude that you ought not kill in MAFIA. Hence her theory would agree with a deontologist theory that you ought not kill in MAFIA even though the Textbook View would consider her theory consequentialist. (Should we consider her theory consequentialist? Again, I briefly postpone answering that question open for now, as my criticism of the Textbook View is consistent with either answer.)

<sup>16</sup> I owe the point that agent-neutral consequentialist theories can posit constraints to Caspar Hare. In an unpublished manuscript, Hare considers an axiological principle like KILLINGS and rejects it as implausible. In its place, he suggests that an improvement would be to place a special importance on the actions in virtue of which two outcomes differ when ranking them from an agent-neutral perspective. He observes that by doing so the consequentialist can make the same extensional claims about which actions are right and wrong as a deontologist.

<sup>17</sup> Since my intention is to provide a counterexample to a key component of the Textbook View, my criticism does not depend on the plausibility of the theory I offer as a counterexample. Still, we might independently wonder how plausible a theory with the claim, KILLINGS, is. I do not think such a theory is correct, but neither do I think it is off the wall. Some people claim that the deontic status of certain actions depends on their causal relations to other actions. In light of this, it is not such a leap to make the teleological claim that an outcome's value depends on the causal relations between its constituent actions. Thanks to Wayne Davis for prompting me to address this point.



Notice that no special work is done by the fact that we are talking about *killings*, as opposed to *lies*, say. The same point could be made for any type of constraint. Very generally, whenever a deontologist posits a constraint against  $\Phi$ -ing, a theorist who holds that you ought to bring about the impersonally best outcome could posit an axiological principle like:

$\Phi$ -ING. If  $\Phi$ -ing,  $\Phi_1$ , would prevent  $\Phi$ -ings,  $\Phi_2$  and  $\Phi_3$ , then an outcome with  $\Phi_1$  is *pro tanto* impersonally worse than an outcome with  $\Phi_2$  and  $\Phi_3$ .

As a result, this theorist can agree extensionally about what you ought to do in cases involving  $\Phi$ -ing with a deontologist who posits a constraint on  $\Phi$ -ing.<sup>18</sup> And the Textbook View will consider this theorist a consequentialist. Thus, we have arrived at the result indicated earlier:

EXTENSIONAL AGREEMENT. Some theories that the Textbook View would consider consequentialist agree extensionally with deontological theories. Specifically, they agree that you ought not kill in *MAFIA*.

This is a significant result. We saw that it is precisely cases like *MAFIA* that the Textbook View takes to separate consequentialists from deontologists extensionally.

If a theory that the Textbook View considers consequentialist agrees extensionally with a deontological theory, what difference is there between these theories?<sup>19</sup> The answer to this question depends on how we individuate moral theories. If we individuate moral theories according to their first-order claims about which actions are right and wrong, then these theories are identical. But I happen to think that we would do better to distinguish moral theories if their explanatory structures differ when they are represented in the theories' canonical formulations. Consider the theory that the Textbook View considers consequentialist. In its canonical formulation, a friend of this theory may state it as a maximizing, teleological theory. This would be because she finds something attractive about the very idea of maximizing and the very idea of a theory that gives an important role to goals. By contrast, consider the deontologist theory. In its canonical formulation, a friend of this theory may state it in terms of deontic claims, rather than teleological claims. She may find this the most illuminating way of portraying what she

<sup>18</sup> This point concerns a simple case where the prevented harms are two further harms of the same type. But the point generalizes to cases that involve more complicated causal dependencies between tokens of action-types. For these more complicated cases, the consequentialist would need to supply other axiological principles. But we have a general recipe for seeing how she would do so in order to generate the same extensional results as a deontologist. According to the Textbook View, the deontologist posits principles with this form:

AGENT-RELATIVE GENERAL CONSTRAINT. Each agent should ensure that she does not  $\Phi$  to prevent a group of others from  $\Omega$ -ing.

We can observe that for every principle with the above form, there is an axiological claim with the following form,

GENERAL AXIOLOGICAL CLAIM. If  $\Phi$ -ing,  $\Phi_1$ , prevents  $\Omega$ -ings,  $\Omega_1, \Omega_2, \dots, \Omega_n$ , then an outcome with  $\Phi_1$  is *pro tanto* impersonally worse than an outcome with  $\Omega_1, \Omega_2, \dots, \Omega_n$ .

A consequentialist can posit principles of the latter form in order to generate the same extensional results as a deontologist who posits principles of the former form.

<sup>19</sup> Thanks to an anonymous reviewer for suggesting that I address this question.

finds attractive about the theory. (Perhaps for her, insofar as claims about goodness are important, these are claims about the good will or good character.) I think there is a reasonable case to be made for distinguishing these theories on the grounds that in their canonical formulations, they appeal to different ethical concepts in their explanations. This is because we are often attracted to theories because of these formulations. But I will not insist on this view here. So I leave it to the reader to decide how she thinks we ought to individuate moral theories. The only point I want to press here is the following. The Textbook View claims that we can separate two classes of theories, the consequentialist and the deontological theories and we can do so, according to an extensional difference between them. And I think it is a mistake to accept this claim.

#### 4 An alternative to the Textbook View

Thus, I suggest that we do away with the Textbook View because both parts of it will not do. The first plank of the Textbook View is the claim,

EXTENSIONAL DISAGREEMENT. Consequentialists and deontologists disagree extensionally. Specifically, they disagree about cases like *MAFIA*.

Against this claim, I have argued that consequentialists can agree extensionally with non-consequentialists about what you ought to do in cases like *MAFIA*. The second plank is the claim,

INTENSIONAL DISAGREEMENT. Consequentialism is an agent-neutral theory. Deontology is an agent-relative theory.

Against this claim, I have argued that there are some agent-neutral deontological theories.

But if we are to throw out the Textbook View, what should we put in its place? Here I think we should take a step back and consider our reason for being interested in the Textbook View in the first place. The reason is to characterize a dispute between two opposing traditions in moral philosophy—the deontological tradition and the consequentialist tradition. The Textbook View aimed to find a division in logical space between agent-relative and agent-neutral theories with which to distinguish these two camps. I have argued that this will not work. One option that is open to us when rejecting the Textbook View is to look for another division in logical space to play this role. I confess that I do not know what this alternative would be. But alternatively, we could give up entirely on the entire enterprise of marking a sharp boundary in logical space, which *possible* moral theories might occupy.<sup>20</sup> Since we are concerned with actual traditions, we should concern

<sup>20</sup> My proposal is inspired by Daniel Jacobsen's claim that Utilitarianism is a movement in the history of ideas. Jacobsen would resist this proposal since he does not think of consequentialism as such a movement, but instead as a "philosophers' term of art, which means whatever philosophers have meant by it over the past half-century or so, when the term was coined and earned its place in the philosophical lexicon." I am sympathetic to Jacobsen's claim, but I think that the fundamental use for the term in the lexicon is to pick out the tradition in moral philosophy that I point to. Interestingly, Jacobsen claims that John Stuart Mill did not endorse an agent-neutral theory. Jacobsen (2008, p. 164).

ourselves primarily with theories that people have actually defended. In doing so, we can set aside imaginary theories that we have to acknowledge people could hold, such as a theory that holds a killing is much worse than a mere death. (Has anyone actually made a serious effort to motivate and defend such a theory?) Instead, we could think of the consequentialist tradition, for example, as beginning in full force with the classical utilitarians. The tradition includes thinkers like G. E. Moore, who held a pluralist theory of the good, so that that beauty, for example, is good as well as happiness. And the tradition ends up with contemporary theorists such as those who hold that the impersonal value of someone's happiness depends on whether she deserves it.<sup>21</sup> What ties these theories together into a tradition is the impact of the thoughts of earlier theorists on later theorists, and the fact that consequently these theories have certain paradigmatic features. These theories have a maximizing structure, they enjoin us to promote agent-neutral ends and they make certain extensional claims about cases like *MAFIA*. If a theory has all of these paradigmatic features, then it is clearly in the consequentialist tradition. If it lacks all these features, then it is clearly not a consequentialist theory. I think that on this basis, we will be able to categorize actual moral theories as we would intuitively like to. But there may be some merely possible theories that have some of these features but not others. An example would be the theory we saw earlier that requires you to bring about the best outcome and endorses *KILLINGS*. This theory is agent-neutral while opposing the paradigmatically consequentialist view of *MAFIA*. Because it has some, but not all, the paradigmatic features of consequentialism, such a theory would be a borderline case of consequentialism: it is indeterminate whether to count it as consequentialist or not. Indeed, that independently seems to me the right view to take of such a theory. I take this to count in favor of an approach that characterizes consequentialism as a tradition of actual theories with paradigmatic features.

**Acknowledgments** For helpful comments and discussions, I would like to thank Wayne Davis, Daniel Halliday, Alejandro Pérez Carballo, Paulina Sliwa, Ekaterina Vavova, the participants in the 2011 Ethics in Society Post-doc Workshop at Stanford University, and an anonymous reviewer for *Philosophical Studies*. I would like to thank Caspar Hare in particular for invaluable discussions and feedback on multiple drafts of this essay.

## References

- Anscombe, G. E. M. (1958). Modern moral philosophy. *Philosophy*, 33, 1–19.
- Bentham, J. (1903). *An Introduction to the Principles of Morals and Legislation*; G. E. Moore, *Principia ethica*, rev. ed., ed. Thomas Baldwin. Cambridge: Cambridge University Press.
- Broome, J. (1991). *Weighing goods*. New York: Blackwell.
- Dreier, J. (1993). The structure of normative theories. *The Monist*, 76, 22–40.
- Feldman, F. (1995). Adjusting utility for justice: A consequentialist reply to the objection from justice. *Philosophy and Phenomenological Research*, 55(3), 567–585.
- Jacobsen, D. (2008). Utilitarianism without consequentialism. *Philosophical Review*, 117(2), 159–191.
- Kagan, S. (1989). *The limits of morality*. Oxford: Clarendon Press.
- McNaughton, D., & Rawling, P. (1991). Agent-relativity and the doing-happening distinction. *Philosophical Studies*, 63(2), 167–185.
- McNaughton, D., & Rawling, P. (1992). Honoring and promoting values. *Ethics*, 102(4), 835–843.

<sup>21</sup> See, for example, Bentham (1903) and Feldman (1995).

- Nozick, R. (1974). *Anarchy state and Utopia*. New York: Basic Books.
- Pettit, P. (2000). Non-consequentialism and universalizability. *The Philosophical Quarterly*, 50(199), 175–190.
- Portmore, D. (2005). Combining teleological ethics with evaluator relativism: A promising result. *Pacific Philosophical Quarterly*, 86, 95–113.
- Ridge, M. (2008). Reasons for action: agent-neutral vs. agent-relative. In: E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall Edition). <http://plato.stanford.edu/archives/fall2008/entries/reasons-agent/>.
- Scheffler, S. (1994). *The rejection of consequentialism*. Oxford: Oxford University Press.
- Schroeder, M. (2007). Teleology, agent-relative value, and good. *Ethics*, 117, 265–295.
- Sen, A. (1982). Rights and agency. *Philosophy and Public Affairs*, 11, 3–39.
- Smith, M. (2009). Two types of consequentialism. *Philosophical Issues*, 19(1), 257–272.
- Thomson, J. J. (1993). Goodness and utilitarianism. *Proceedings and Addresses of the American Philosophical Association*, 67, 145–159.